



Stress cellulaire et modulation de l'activité des cytidines désaminases APOBEC3

Mohamed Salah Bouzidi

► To cite this version:

Mohamed Salah Bouzidi. Stress cellulaire et modulation de l'activité des cytidines désaminases APOBEC3. Virologie. Université Pierre et Marie Curie - Paris VI, 2015. Français. NNT : 2015PA066597 . tel-01334908

HAL Id: tel-01334908

<https://theses.hal.science/tel-01334908>

Submitted on 21 Jun 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Pierre et Marie Curie

Ecole doctorale Complexité du Vivant-ED515

Unité de Rétrovirologie Moléculaire / Institut Pasteur

Stress cellulaire et modulation de l'activité des cytidines désaminases APOBEC3

Par Mohamed Salah BOUZIDI

Thèse de doctorat de Virologie

Dirigée par Messieurs Jean-Pierre VARTANIAN et Simon WAIN-HOBSON

Présentée et soutenue publiquement le 29 Septembre 2015

Devant un jury composé de :

Monsieur Vincent MARÉCHAL

Président

Monsieur Andrea CIMARELLI

Rapporteur

Monsieur Jean-Christophe PAGÈS

Rapporteur

Monsieur Christophe TERZIAN

Examineur

Monsieur Jean-Pierre VARTANIAN

Directeur de thèse

Monsieur Simon WAIN-HOBSON

Co-Directeur de thèse



Except where otherwise noted, this work is licensed under
<http://creativecommons.org/licenses/by-nc-nd/3.0/>

Université Pierre et Marie Curie

Ecole doctorale Complexité du Vivant-ED515

Unité de Rétrovirologie Moléculaire / Institut Pasteur

Stress cellulaire et modulation de l'activité des cytidines désaminases APOBEC3

Par Mohamed Salah BOUZIDI

Thèse de doctorat de Virologie

Dirigée par Messieurs Jean-Pierre VARTANIAN et Simon WAIN-HOBSON

Présentée et soutenue publiquement le 29 Septembre 2015

Devant un jury composé de :

Monsieur Vincent MARÉCHAL

Président

Monsieur Andrea CIMARELLI

Rapporteur

Monsieur Jean-Christophe PAGÈS

Rapporteur

Monsieur Christophe TERZIAN

Examineur

Monsieur Jean-Pierre VARTANIAN

Directeur de thèse

Monsieur Simon WAIN-HOBSON

Co-Directeur de thèse



Except where otherwise noted, this work is licensed under
<http://creativecommons.org/licenses/by-nc-nd/3.0/>

Dédicace

Remerciements

RESUME.....	I
TABLE DES FIGURES.....	II
TABLE DES ABREVIATIONS.....	V
INTRODUCTION.....	1
I. LES MUTATIONS GENETIQUES ET LEURS SOURCES	1
A. Les différents types de mutations génétiques	1
1. Les substitutions	1
2. Les insertions et les délétions.....	2
B. Les sources des mutations génétiques	2
1. Les agents mutagènes physiques.....	2
2. Les agents mutagènes chimiques	3
3. Les désaminations	4
a. Désamination des bases non méthylées.....	4
b. Désamination des bases méthylées.....	4
4. Les cytidines désaminases cellulaires	5
a. CDA.....	6
b. APOBEC/AID	6
i. APOBEC 1	7
ii. APOBEC 2	9
iii. APOBEC 4.....	9
iv. AID.....	10
v. APOBEC3	11
α. APOBEC3A	12
β. APOBEC3B	14
γ. APOBEC3C.....	14
δ. APOBEC3DE	15
ε. APOBEC3F	15
ζ. APOBEC3G.....	16
η. APOBEC3H	17
vi. Structure des A3	18
vii. Evolution des cytidines désaminases APOBEC/AID	18
viii. Signature mutationnelle des AID/APOBEC	20
II. APOBEC ET VIRUS.....	21
A. Historique	21
B. Réponse immunitaire et APOBEC	22
C. APOBEC et Rétrovirus « complexes ».....	23
1. Le Virus de l'Immunodéficience Humaine	24
2. Le Virus T-Lymphotropique Humain	25
3. Le virus Foamy.....	26
4. Les Rétrovirus endogènes	27
D. APOBEC3 et Virus à ADN	29
1. Les virus à ADN double brin	29
a. Les papillomavirus humains.....	29
b. Les Herpesvirus	31
2. APOBEC 3 et virus à ADN simple brin.....	31
3. Le virus de l'Hépatite B	32
III. ROLE DES APOBEC DANS LE CATABOLISME ET LE CANCER.....	36
A. A3 et catabolisme de l'ADN.....	37

1. Les Hypermutations	37
2. Les censeurs de l'ADN cytoplasmique	39
a. La voie du Toll-Like Receptor 9	39
b. La voie c-GAS (Cyclic GMP-AMP synthase)	40
c. La protéine STING	40
d. La voie de RIG-I et de l'ARN polymérase III.....	42
3. L'Uracile-ADN glycosylase.....	43
4. Les hypomutations	44
<i>B. La tumorigénèse.....</i>	<i>45</i>
1. Nature des mutations dans les cancers	46
2. Implication d'AID dans la tumorigénèse	47
3. Implication de A1 et de A2 dans la tumorigénèse.....	49
4. Implication des A3 dans le cancer.....	49
a. Le Rôle d'A3B dans l'émergence des cancers.	49
b. Impact de la délétion d'A3B sur la carcinogénèse	50
PROBLEMATIQUES ET OBJECTIFS DES TRAVAUX DE THESE.	54
MATERIELS ET METHODES	56
I. RESULTAT 1	57
<i>A. Patients</i>	<i>57</i>
<i>B. Lignées cellulaires et cultures</i>	<i>57</i>
<i>C. Plasmides.....</i>	<i>58</i>
<i>D. Extraction protéique et purification de co-affinité.....</i>	<i>58</i>
<i>E. Western Blot.....</i>	<i>59</i>
<i>F. Extraction d'ADN, PCR et 3DPCR.</i>	<i>59</i>
<i>F. PCR quantitative.....</i>	<i>60</i>
<i>G. Microscopie confocale.....</i>	<i>60</i>
II. RESULTAT 2	62
<i>A. Plasmides</i>	<i>62</i>
<i>B. Culture cellulaire.....</i>	<i>62</i>
<i>C. Transfections</i>	<i>62</i>
<i>D. Western blot.....</i>	<i>63</i>
<i>E. Immunofluorescence</i>	<i>63</i>
<i>F. Mesure de désamination in vitro</i>	<i>63</i>
<i>G. 3DPCR.....</i>	<i>64</i>
<i>H. Analyse par cytométrie en flux des DSB.....</i>	<i>64</i>
<i>I. RNAseq sur les échantillons de sperme</i>	<i>64</i>
<i>J. Analyse structurale.....</i>	<i>65</i>
III- RESULTAT 3	66
<i>A. Cellules et échantillons</i>	<i>66</i>
<i>B. PCR et 3DPCR.....</i>	<i>66</i>
IV-RESULTATS 4	68
<i>A. Réactifs.....</i>	<i>68</i>
<i>B. Plasmides</i>	<i>68</i>
<i>C. Culture cellulaire et transfection.....</i>	<i>68</i>
<i>D. PBMC et isolation des lymphocytes T CD4+.....</i>	<i>69</i>
<i>E. Tri cellulaire</i>	<i>69</i>
<i>F. Relargage du cytochrome c</i>	<i>69</i>
<i>G. Analyse de l'apoptose par cytométrie.</i>	<i>70</i>
<i>H. Analyse des cassures de l'ADN double brin par cytométrie en flux.</i>	<i>70</i>

TABLE DES MATIERES

<i>I. PCR/3DPCR</i>	70
<i>J. PCR quantitative</i>	71
<i>K. Western Blot</i>	71
<i>L. Co-immunoprecipitation ARN/protéine.</i>	71
RESULTATS	73
I. RESULTAT 1 : REGULATION DE L'EDITING DU VIRUS DE L'HEPATITE B PAR HETERODIMERISATION DES CYTIDINES DESAMINASES APOBEC3	74
<i>A. Introduction</i>	74
<i>B. Résultats</i>	76
1) APOBEC3DE forme des dimères avec APOBEC3F et APOBEC3G	76
2) APOBEC3DE inhibe l'editing du VHB par APOBEC3F et APOBEC3G	77
3) APOBEC3DE promeut la réplication du VHB	78
4) APOBEC3DEC320Y hyperédite le génome du VHB.	79
<i>C. Conclusion</i>	80
II. RESULTAT 2 : BASES MOLECULAIRES DU PHENOTYPE ATTENUÉ DE L'ENZYME MUTATRICE DE L'ADN APOBEC3B	81
<i>A. Introduction</i>	81
<i>B. Résultats</i>	81
1. Etude de l'impact des deux domaines d'A3B.	81
2. Les déterminants moléculaires de l'atténuation d'A3Bc	83
3. L'atténuation d'A3Bc est conservée dans l'évolution	84
<i>C. Conclusion</i>	85
III. RESULTAT 3 : IDENTIFICATION ERRONÉE D'ADN CHROMOSOMIQUE ÉDITÉ PAR LES APOBEC-3 DANS LES GÉNOMES DE CANCERS.	86
<i>A. Introduction</i>	86
<i>B. Résultats</i>	87
1. Détermination du taux d'erreur lors d'une 3DPCR	87
2. Détermination du taux d'erreur lors d'une 3DPCR avec un tampon <i>long range</i>	88
3. La 3DPCR permet de mettre en évidence des séquences faiblement éditées par les A3	88
4. La 3DPCR induit une complexification des séquences mutées par recombinaison	90
<i>C. Conclusions</i>	91
IV. RESULTAT 4 : L'INDUCTION DE LA SUREXPRESSION DE LA CYTIDINE DESAMINASE APOBEC3A PAR L'ADN CYTOPLASMIQUE ENDOGENE PROVOQUE DES DOMMAGES DE L'ADN CHROMOSOMIQUE	93
<i>A. Introduction</i>	93
<i>B. Résultats</i>	94
2. <i>Editing</i> de l'ADNcymt dans les lymphocytes T CD4 ⁺ primaires	95
3. L'etoposide et l'actinomycine D induisent l'apoptose et la surexpression des A3	96
4. ADNmt circulant dans les cas de sepsis.	97
5. L'Uracile dans l'ADN cytosolique ne constitue pas un signal de danger	98
6. L'ADN <i>MT-COI</i> active la réponse interféron conduisant à l'expression des A3	99
7. L'ADN double brin est transcrit par l'ARN polymérase III, ce qui déclenche la voie de signalisation RIG-I	100
<i>C. Conclusion</i>	102
DISCUSSION	104
A. A3DE EST PERMET-ELLE AU VHB D'ÉCHAPPER À LA RESTRICTION PAR A3G ?	105
B. QUEL EST LE RÔLE PHYSIOLOGIQUE D'A3DE ?	107
C. UN RÔLE PHYSIOLOGIQUE POUR A3A ?	109

TABLE DES MATIERES

D.A3A, UN FACTEUR AGGRAVANT DANS LES MALADIES LIEES AU CATABOLISME DE L'ADN MITOCHONDRIAL ?	111
E. QUELLE EST LA CONTRIBUTION RESPECTIVE D'A3A ET D'A3B DANS LA CARCINOGENESE ?	113
BIBLIOGRAPHIE	117
ANNEXES	136

Résumé

Les protéines APOBEC3 (A3A-A3H) catalysent la désamination des cytidines (C) présentes sur l'ADN simple brin en thymidine (T).

Cette activité cytidines désaminase a initialement été décrite comme impliquée dans la restriction des rétrovirus et de certains virus à ADN par leur capacité à induire de nombreuses mutations C->T, ou hypermutations, sur les génomes viraux. Il apparaît néanmoins que leur activité n'est pas restreinte aux génomes viraux et que certaines A3 peuvent induire des mutations sur l'ADN mitochondrial (A3A, C, F, G et H) et nucléaire (A3A et A3B). Ainsi, l'impact somatique des A3 est désormais établi dans la formation de certains cancers, dont la majorité des mutations, portent signatures des APOBEC3.

Aux vues de ces observations, nous nous sommes intéressés à la façon dont sont régulées ces enzymes dans le contexte du stress cellulaire viro-induit ou endogène.

La première partie de nos travaux a porté sur la protéine A3DE, seul membre de la famille APOBEC3 ne possédant pas d'activité cytidine désaminase. De façon intéressante, il apparaît qu'A3DE est surexprimée dans les cirrhoses infectées par le VHB, VHC ou co-infectées par le VHC et le VHB. Nous avons pu mettre en évidence qu'A3DE interagit et module l'activité d'A3F et d'A3G, deux cytidines désaminases exprimées dans le foie et impliquées dans la restriction du VHB.

Dans un second temps, nous nous sommes intéressés à la caractérisation du potentiel génotoxique de la protéine A3B. Cette protéine, de par sa localisation strictement nucléaire, constitue la seule A3 à double domaine n'interagissant pas avec A3DE. Contrairement à A3A, A3B est faiblement active sur l'ADN nucléaire et n'induit pas de cassures de l'ADN double brin. Nous avons pu mettre en évidence par mutagenèse les régions de la protéine impliquées dans l'atténuation de la génotoxicité d'A3B par rapport à A3A et que cette atténuation est conservée chez les primates. Enfin, nous avons étudié le rôle et la régulation d'A3A dans le catabolisme. Nous avons mis en évidence que l'ADN mitochondrial cytoplasmique (ADNcymt) active la voie RIG-I/ARN polymérase III ce qui a pour effet d'induire la production d'IFN qui va activer l'expression d'A3A. A3A va ainsi jouer un rôle dans le catabolisme de l'ADNcymt et contribue à l'élimination de cette source de stress cellulaire, mais occasionnant par la même des dommages sur l'ADN nucléaire.

Les A3 sont des enzymes fondamentales de la défense immunitaire innée et du catabolisme de l'ADN. Nous montrons qu'A3DE a pour fonction de moduler l'activité d'A3F et d'A3G tandis qu'A3B, possède un phénotype atténué chez tous les primates et s'avère moins génotoxique que'A3A. Cette dernière participe à la dégradation de l'ADN cytoplasmique, limitant ainsi l'inflammation. Néanmoins, A3A peut s'avérer dangereuse pour l'intégrité génomique et contribuer à l'émergence de cancers, notamment en cas d'inflammation chronique.

Table des figures

Figure 1 : Schéma des différents types de mutations possibles sur l'ADN.

Figure 2 : Différentes réactions de désaminations possibles sur l'ADN.

Figure 3 : Structure cristallographique de la cytidine désaminase CDA.

Figure 4 : Les différents membres de la famille APOBEC.

Figure 5 : Mécanismes de diversification des gènes d'immunoglobulines par AID.

Figure 6 : Organisation et localisation cellulaire des APOBEC3.

Figure 7 Représentation de la structure prédite de hA3DE classique et hA3DE C320Y.

Figure 8 : Transcriptome des APOBEC3 sur des tissus cirrhotiques de patients sains et de patients infectés par VHB, le VHC ou co infectés VHB et VHC.

Figure 9 : Représentation schématique des A3 en fonction de leur type de domaine catalytique Z1, Z2 ou Z3.

Figure 10 : Modèle de la structure d'A3A obtenue par NMR

Figure 11 : Evolution des APOBEC/AID chez les vertébrés.

Figure 12 : Contexte dinucléotidique de désamination des différentes A3.

Figure 13 : Schéma du mécanisme d'induction de l'expression des A3 par l'IFN de type I.

Figure 14 : Schéma de la restriction du VIH par A3G.

Figure 15 : Mécanisme de dégradation d'A3G par la protéine Vif.

Figure 16 : Les rétrovirus endogènes.

Figure 17 : Principe de la 3D-PCR.

Figure 18 : Mécanisme d'induction de l'expression de gènes pro inflammatoire par l'ADN CpG par l'intermédiaire du TLR9.

Figure 19 : Mécanisme d'induction de la réponse immunitaire par l'intermédiaire de cGAS et STING.

Figure 20 : Mécanisme d'induction de la réponse immunitaire par l'ARN double brin par l'intermédiaire de RIG-I.

Figure 21 : Spectre des mutations dans les cancers humaines.

Figure 22 : Nature et signature des mutations C->T dans le cancer du sein PD4120.

Figure 23 : Schéma du locus A3 normal ou ΔA3B portant une délétion de 29.5kb.

Figure 24 : A3DE dimérise avec elle-même, A3F et A3G.

Figure 25 : Microscopie confocale des A3DE avec d'autres protéines A3.

Figure 26 : A3DEC320Y dimérise et colocalise avec lui-même, A3DE, A3F et A3G.

Figure 27 : A3DE protège l'ADN du VHB contre la désamination par les A3 de manière dose-dépendante.

Figure 28 : L'interaction d'A3DE avec A3F et A3G entraîne leur inhibition et l'augmentation des niveaux d'ADN de VHB.

Figure 29 : A3DE_{C320} n'est pas fonctionnelle mais peut acquérir une activité par la mutation C320Y.

Figure 30 : Analyse par FACS des cellules positives pour γ H2AX et analyse par western blot de l'expression des chimères A3 taggées V5.

Figure 31 : Chimères A3A/A3B.

Figure 32 : Représentation schématique d'A3BnA et A3Bc et localisation cellulaire des A3.

Figure 33 : Analyse par FACS des cellules HeLa γ H2AX positives et représentation de la structure d'A3A.

Figure 34 : Alignement des séquences d'A3A et A3Bc d'Homme et de macaque rhésus et localisation cellulaire des A3 de macaque rhésus.

Figure 35 : Evolution du domaine Z1 d'A3A et d'A3B et point Isoélectrique théorique des différents domaines des A3 de primate.

Figure 36 : Transcriptome des cytidines désaminases humaines dans le sperme mature humain purifié.

Figure 37 : Variation du bruit de fond selon la position sur le bloc lors d'une 3DPCR avec la Taq polymérase.

Figure 38 : Contexte dinucléotidique 3' des mutations bruit de fond de la 3DPCR avec un tampon standard ou long range et une Taq polymérase.

Figure 39 : Variation du bruit de fond selon la position sur le bloc lors d'une 3DPCR avec la Taq polymérase et le tampon long range.

Figure 40 : Variation du bruit de fond selon la position sur le bloc lors d'une 3DPCR avec la Pfu polymérase.

Figure 41 : 3DPCR sur des produits PCR de CTNNB1 issus de foie sains ou cirrhotique.

Figure 42 : Séquences CTNNB1 hypermutées par les A3 purifiées de lymphocytes T CD4 humains.

Figure 43 : Editing massif de l'ADNcymt par les A3 dans une cellule.

Figure 44 : Editing de la séquence MT-COI dans une cellule P2.

Figure 45 : Editing de MT-COI dans les lymphocytes T CD4⁺ provenant du donneur 1.

Figure 46 : Editing de MT-COI un lymphocyte T CD4⁺ provenant du donneur 2.

TABLE DES FIGURES

Figure 47 : L'etoposide et l'actinomycine D induisent l'expression des A3 et l'editing de l'ADN MT-COI dans les cellules P2.

Figure 48 : Editing de MT-COI chez les patients sains et atteints de sepsis.

Figure 49 : L'ADN contenant de l'uracile ne constitue pas un nouveau signal de danger.

Figure 50 : Transcriptome des A3 dans les cellules THP-1 transfectées par de l'ADN.

Figure 51 : La transfection d'ADN MT-COI active la production d'IFN α et β dans les PBMC.

Figure 52 : L'ADN transfecté induit des cassures de l'ADN double brin.

Figure 53 : L'ADN double brin transfecté active la voie de signalisation RIG-I via l'ARN polymérase III.

Figure 54 : Voie de signalisation détectant l'ADN cytosolique et déclenchant la réponse immunitaire.

Tableau 1 : Analyse de l'allèle présent d'APOBEC3DE chez des patients et différentes lignées cellulaires.

Tableau 2 : Liste des amorces employées pour les constructions des chimères A3A/A3B pour le système pcDNA3.1 TOPO TA V5 cloning.

Tableau 3 : Liste des amorces employées pour la mutagenèse dirigée d'A3B.

Tableau 4 : Compendium des oligonucléotides utilisés dans cette étude.

Table des abréviations

3DPCR: Differential DNA denaturation polymerase chain reaction

A: APOBEC: apolipoprotein B mRNA editing¹ complex

AAV : Adeno associated virus : Virus Adéno-Associé

ACF : APOBEC-1 complementation factor

ADN: acide désoxyribonucléique

ADN: acide désoxyribonucléique complémentaire

ADNmt : ADN mitochondrial

ADNcymt : ADN mitochondrial cytoplasmique

AID: activation induced deaminase protein

AICDA: activation induced deaminase gene

ApoB: apolipoprotein B

ARNm: acide ribonucléique messagé

ARNpol III : ARN polymérase III

CAEV : Caprine Arthritis Encephalitis Virus : Virus de l'arthrite/encéphalite caprine

cccDNA : covalently closed circular DNA

CDA: Cytidine deaminase

CHC: Carcinome Hépatocellulaire

c-GAS : Cyclic GMP-AMP synthase

C-Terminal : Carboxy-terminal

db: double brin:

DSB: Double Strand Break (Cassures de l'ADN double brin)

E.coli: Escherichia coli

EBV: Epstein-Barr virus

ERV: rétrovirus endogène

FACS: Fluorescence-activated cell sorter (cytomètre de flux)

GST: Glutathione S-transferase

HPV: Human Papillomavirus

HRP: Horseradish peroxidase

HSV: Herpes simplex virus

HTLV: Human T cell leukemia virus

¹ Le terme *editing* sera employé pour désigner les mutations induites par une APOBEC tout au long de ce manuscrit.

IFN: interféron
Ig: Immunoglobuline
IL: interleukine
ISG: Interferon Stimulated Gene:
IRF: Interferon Regulatory Factor
kb: kilobases
Da: Dalton
LINE: long interspread element
LTR: Long terminal repeat
Mb: méga base
meC: Méthyle cytidine
MMR: mismatch repair
MLV: Murine leukemia virus
NES: Nuclear export signal
NFκB: nuclear factor kappa B
N-Ter : Amino terminal
pdb: paire de base
PBMC: peripheral blood mononuclear cell (cellules mononuclées du sang périphérique)
PCR: polymerase chain reaction
PGA : Peptidoglycane
PHA: Phytohemagglutinin
PMA : phorbol12-myristate 13-acetate
ROS : Reactive oxygen species : Espèces réactives de l'oxygène
RT-PCR: reverse transcriptase polymérase chain reaction
siARN : ARN interférant
SIV: Simian Immunodeficiency virus (Virus de l'Immunodéficience Simienne)
Td= Température de dénaturation
Tm= Température of melting = température de fusion
TLR: Toll like receptor
TNF: tumor necrosis factor
TRB: tribble
UGI: uracyl N glycosilase inhibitor
UNG: uracyl N glycosilase
UTR : Untranslated Region : Région non traduite

ABBREVIATIONS

UV: ultra-violet

vif: viral infectivity factor

VHB: Virus de l'Hépatite B

VHC: Virus de l'Hépatite C

VIH: Virus de l'Immunodéficience Humaine

INTRODUCTION

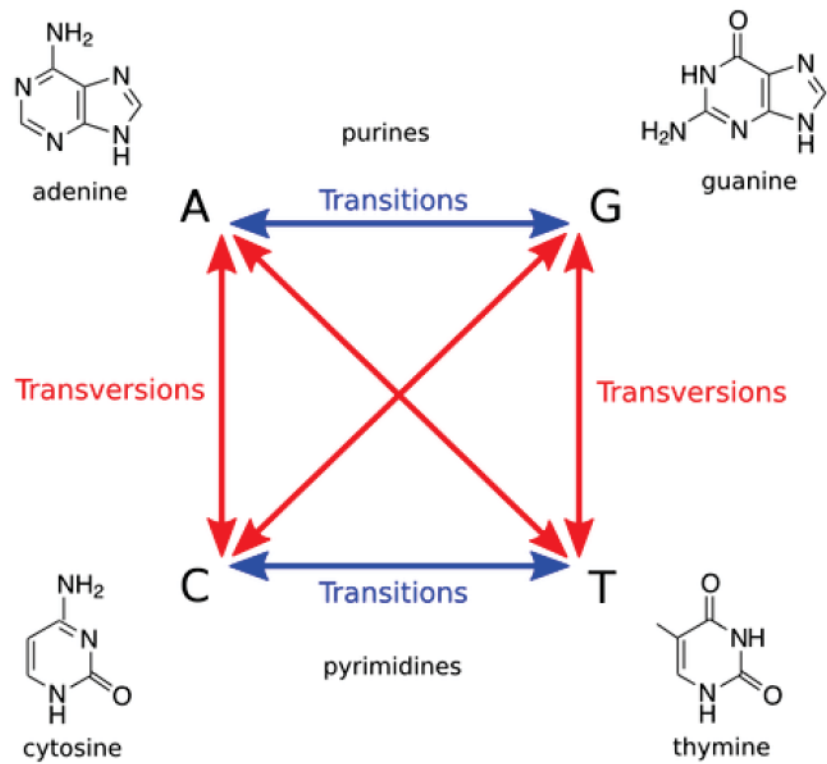


Figure 1 : Schéma des différents types de mutations possibles sur l'ADN.

I. Les mutations génétiques et leurs sources

A. Les différents types de mutations génétiques

Le phénomène de mutation se définit par la modification de l'information génétique dans le génome d'un organisme. Il existe plusieurs types de mutations : les mutations par substitution, les insertions et les délétions.

1. Les substitutions

Les mutations par substitution sont le remplacement d'une base par une autre. Les bases azotées sont classées en deux grandes familles : les bases puriques, qui sont des molécules azotées hétérocycliques, constituées d'un cycle pyrimidine fusionné à un cycle imidazole (Adénine et Guanine), et les bases pyrimidiques, dont la structure azotée hétérocyclique aromatique est voisine de la pyridine (thymine et cytosine).

On appelle transitions les substitutions d'une base azotée par une autre de même famille (pyrimidine à pyrimidine ou purine à purine) et transversion le passage d'une base azotée d'une famille à une base azotée d'une autre famille (pyrimidine à purine ou purine à pyrimidine) (Figure 1). Ces mutations sont qualifiées de mutations faux sens lorsqu'elles entraînent un changement de la séquence en acides aminés au niveau protéique. Elles sont qualifiées de mutations non-sens lorsqu'elles entraînent la formation d'un codon stop dans la séquence protéique. Enfin, les mutations sont dites silencieuses dans le cas où la mutation n'affecte pas la séquence en acides aminés du fait de la redondance du code génétique, ou alors dans le cas où la mutation a lieu dans une région non transcrite du génome ou sur un intron. Néanmoins, il arrive que des mutations dites silencieuses induisent un changement de phénotype, dans le cas où la mutation a lieu sur un site d'épissage par exemple. En effet, les sites d'épissage sont les lieux d'excision des introns. Une mutation à ce niveau peut donc empêcher ce mécanisme et aboutir à la traduction d'un ARN dont un ou plusieurs introns demeurent, et par conséquent à l'expression d'une protéine potentiellement non fonctionnelle.

2. Les insertions et les délétions

Les insertions et les délétions (indel) constituent à elles deux le second grand type de mutations. Ces mutations sont dites « décalantes » lorsqu'elles ne sont pas des insertions ou des délétions d'un nombre de nucléotide multiple de 3, et qu'elles entraînent un décalage du cadre de lecture lors de la transcription. Ceci peut avoir comme conséquence la production d'une protéine tronquée du fait de l'apparition d'un codon stop prématuré. Les indels peuvent avoir des tailles très variables, allant d'une base à dix milles bases. Les indels sont un mécanisme important dans l'évolution et la diversité génétique : on estime qu'ils constituent la deuxième source de variabilité derrière les substitutions (Montgomery et al., 2013).

B. Les sources des mutations génétiques

Ces mutations peuvent avoir plusieurs sources. L'une des principales sources de mutations sont les erreurs de l'ADN polymérase. Le taux d'erreur de l'ADN polymérase chez l'Homme est d'environ 10^{-8} mutations par paire de bases (pdb)(Kunkel, 2004). La machinerie de réparation cellulaire corrige environ 99% de ces erreurs. Ceci aboutit à un taux d'erreur de 10^{-10} mutations par pdb. La taille du génome humain haploïde étant d'environ 3400 Mpdb, il y a donc environ 0,34 mutations par division pour une cellule de la lignée germinale et 0,68 mutations par cellule de la lignée somatique (Drake, Charlesworth, Charlesworth, & Crow, 1998).

Les mutations peuvent aussi être le résultat d'une source exogène. Ces facteurs dits « environnementaux » vont par conséquent avoir un impact différent sur les organes et tissus selon leur exposition. Ces facteurs environnementaux peuvent être classés en 2 catégories selon leur nature : physique ou chimique.

1. Les agents mutagènes physiques

La majorité des mutagènes physiques sont les radiations. C'est le cas par exemple des ultraviolets qui vont fournir l'énergie nécessaire à la formation de dimères de thymine dans l'ADN (Goodsell, 2001). Lorsque ces dimères ne sont pas réparés par la cellule, ils vont

empêcher une bonne lecture du brin matrice par l'ADN polymérase et entraîner un taux d'erreurs accru lors de la réplication (Goodsell, 2001). La majorité des mutations observées seront des mutations T->C. Les radiations touchent souvent les tissus et organes superficiels (peau, yeux, etc...), c'est donc dans ces tissus que l'on va retrouver le plus de mutations de ce type. Parmi les mutagènes physiques, on retrouve aussi les rayons ionisants (rayons alpha, gamma, X, etc..) causant des cassures de l'ADN double brin (DSB) (Santivasi & Xia, 2014)

2. Les agents mutagènes chimiques

Parmi les mutateurs chimiques, on retrouve le benzopyrène, présent dans les gaz d'échappement ou la fumée de cigarette. Le benzopyrene est un intercalant de l'ADN. Sa présence va perturber la structure de l'ADN et empêcher une bonne lecture du brin matrice par la polymérase (Baird, Hooven, & Mahadevan, 2005). L'aflatoxine, un autre mutagène chimique, induit des mutations de façon similaire au benzopyrène. Les dérivés réactifs de l'oxygène (ROS) telle que les radicaux libres, les peroxydes et ions oxygénés sont de puissants mutateurs. Ceux-ci entraînent la formation de sites abasiques entraînant des cassures de l'ADN simple brin, mais aussi des modifications de guanine en 8-oxoguanine. Les 8-oxoguanines sont mal lues par l'ADN polymérase et appariées à des adénines, induisant ainsi des transversions G->T (Jena, 2012). On estime à 20 000 par jour le nombre de mutations induites par les ROS dans une cellule.

Toutes les mutations ne sont pas le résultat de l'action d'un facteur chimique ou physique. La modification d'une fonction particulière des bases azotées, la fonction amine, peut avoir lieu spontanément ou suite à l'intervention de facteurs cellulaires. Cette fonction étant impliquée dans la formation de liaison hydrogène, ceci va impacter la façon dont la base désaminée, c'est-à-dire privée de sa fonction amine, va s'apparier avec les autres bases et aboutir à des mutations.

Enfin, certaines mutations peuvent apparaître spontanément ou par l'activité de facteurs cellulaire : c'est le cas de la désamination.

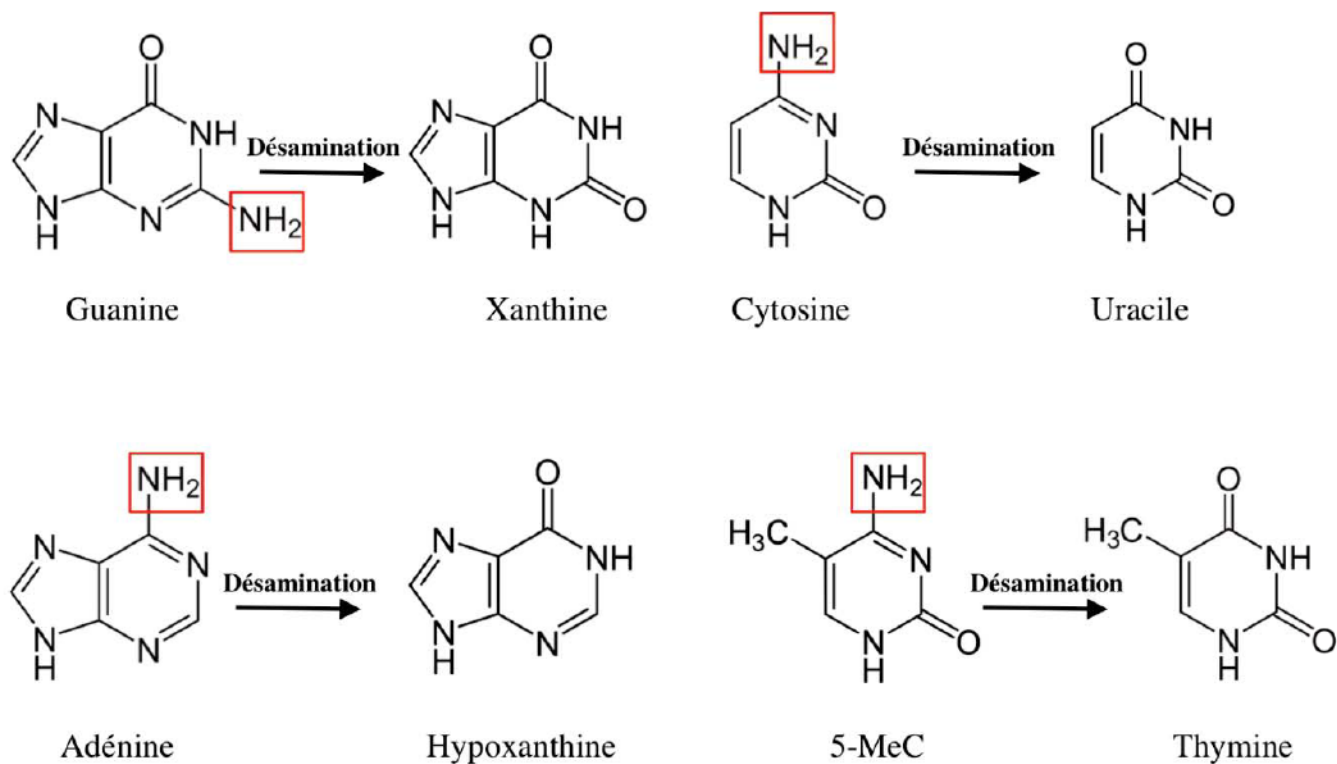


Figure 2 : différentes réactions de désaminations possibles sur l'ADN. La fonction amine subissant la désamination est indiquée en rouge

3. Les désaminations

a. Désamination des bases non méthylées

La désamination est la réaction d'hydrolyse d'un groupement amine d'une molécule entraînant la libération d'une molécule d'ammoniac (NH_3). C'est une réaction qui est à la base de la dégradation des acides aminés. Néanmoins, elle est aussi importante au niveau de l'ADN. En effet, toutes les bases azotées, à l'exception de la thymidine, possèdent un groupement amine. Le clivage de ce groupement va entraîner des mutations en absence de correction par le système de réparation.

La désamination d'une adénine entraîne la formation d'une inosine (Figure 2). L'inosine est notamment retrouvée dans les séquences d'ARN de transfert, permettant à ceux-ci de s'apparier avec plusieurs codons synonymes. Dans l'ADN, l'inosine s'apparie avec la cytosine aboutissant après réplication à des transitions A->G (Kow, 2002).

La xanthine est le résultat de la désamination d'une guanine (Figure 2). La xanthine se lie préférentiellement à la thymine et ceci aboutit donc à des transitions G->A lorsque celles-ci ne sont pas corrigées par la machinerie cellulaire (Kow, 2002).

La désamination des cytosines va entraîner la formation de résidus uraciles dans l'ADN (Figure 2). Ces résidus peuvent être corrigés par le système de réparation par excision de bases (BER) (ref). Ce système fait intervenir une glycosylase particulière : l'Uracile DNA Glycosylase (UNG). Les mutations qui échappent au BER vont aboutir à l'apparition des transitions C->T (Kow, 2002)

b. Désamination des bases méthylées

La méthylation est une modification épigénétique réversible de l'ADN, qui consiste en l'ajout d'une fonction méthyle ($-\text{CH}_3$) sur le carbone 5 de la base azotée, sur les résidus adénine (chez les bactéries) et cytosine. La méthylation des cytosines chez l'Homme est assurée par des enzymes appelées les DNA methyl transférases. Cette famille d'enzymes se compose de 4 membres chez l'Homme (Dnmt 1 à 4) en présence de S-adenosyl-méthionine (SAM) (Kulis & Esteller, 2010).

Les méthylations ont principalement lieu sur les cytosines des dinucléotides CpG, bien que la méthylation des cytosines soit observée dans d'autres configurations dinucléotidiques dans les cellules souches et les cellules embryonnaires. Les dinucléotides CpG sont souvent regroupés dans des régions appelées îlots CpG. Ces îlots sont retrouvés notamment au niveau des promoteurs et du premier exon d'un gène. En effet, la méthylation est un élément de contrôle épigénétique de la transcription de gènes : les groupements méthyles sont reconnus par les méthyl CpG binding protein (MBD) ce qui permet le recrutement d'histones déacétylases (HDAC). Ceci entraîne la compaction de la chromatine inhibant ainsi l'expression des gènes. Des défauts de méthylation peuvent être associés à certaines pathologies : réactivation de virus oncogènes latents, mélanome, cancer colorectal (Kulis & Esteller, 2010). Les cytidines méthylées sont aussi sensibles à la déamination, qu'elle soit spontanée (avec un taux de 10^{-13} environ) ou le produit de l'activité de facteurs cellulaires dont le résultat va être une transition cytidine méthylée vers thymine (Shen, Rideout, & Jones, 1994) (Figure 2). Ceci va donc avoir un impact sur la régulation de l'expression des gènes et éventuellement mener à l'apparition de cancers. On estime à 1% le nombre de cytosine méthylées dans le génome et celles-ci comptent pour 30% des mutations germinales et somatiques.

De nombreux génomes de cellules cancéreuses portent la marque de cytidines désaminées (mutations C->T) mais aussi de 5meC mutées (Nik-Zainal et al., 2012) (Greenman et al., 2007). Par ailleurs, il a été montré que certaines cytidines désaminases sont impliquées dans les défenses antivirales et la carcinogénèse (Burns, Lackey, et al., 2013) . Pour ces raisons, nous nous sommes concentrés dans ce travail sur les cytidines désaminases.

4. Les cytidines désaminases cellulaires

Les cytidines désaminases sont des hydrolases qui catalysent la déamination d'un résidu cytidine en uracile. Il en existe deux grands types : la cytidine désaminase Humaine (CDA), impliquée dans le maintien du taux de pyrimidine intracellulaire et celles de la famille AID/APOBEC dont les cibles sont les acides nucléiques.

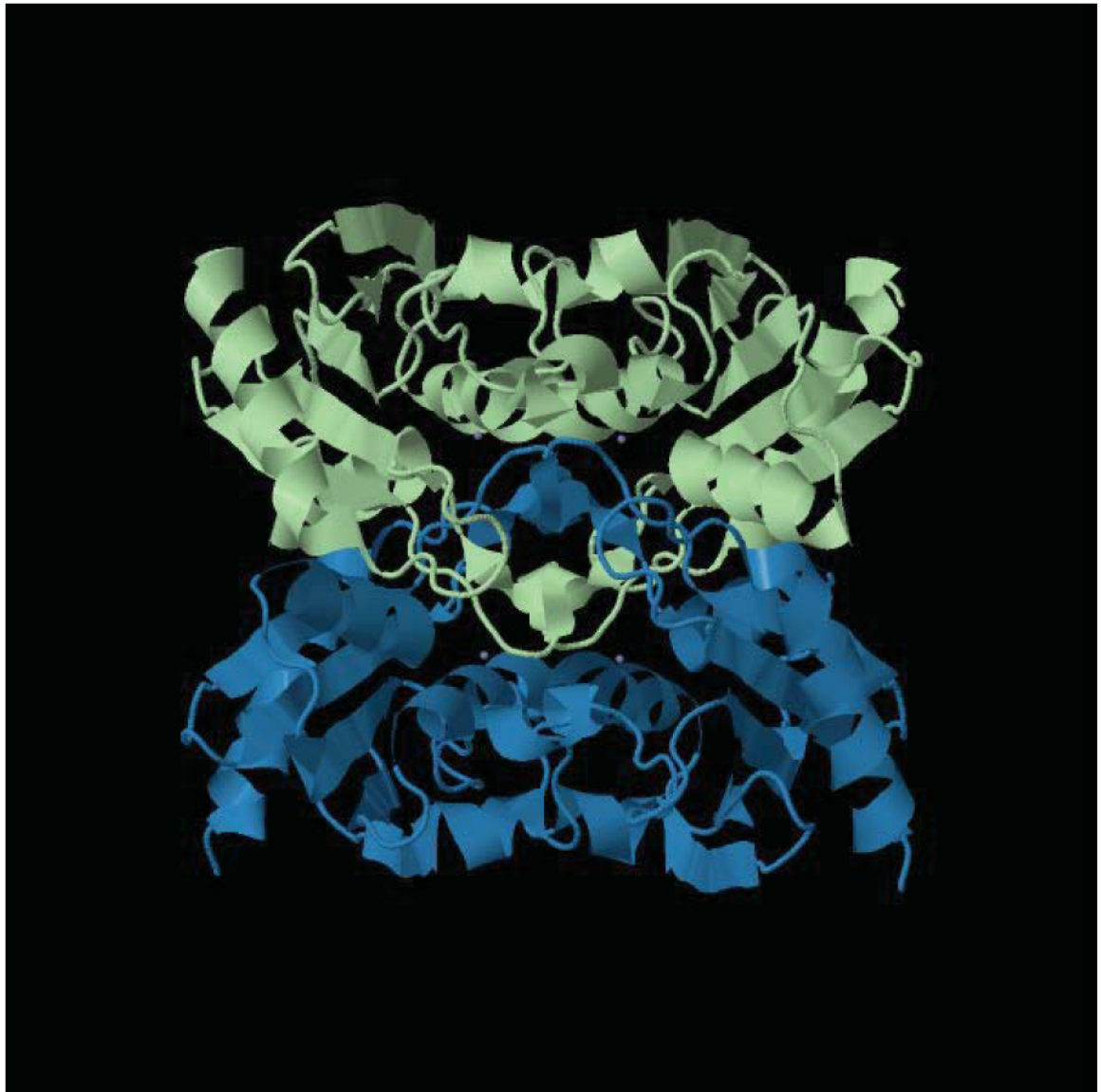


Figure 3 : Structure cristallographique de la cytidine désaminase CDA. La protéine est représentée sous forme d'un homotétramère pour lesquels les deux homodimères symétriques sont colorés en vert et en bleu. (Adapté de Chung et al. 2005)

a. CDA

La cytidine désaminase CDA est codée par le gène *CDA* localisé sur le chromosome 1 chez l'Homme (Saccone et al., 1994). Elle est retrouvée aussi bien chez les eucaryotes que chez les procaryotes ce qui témoigne d'une émergence ancienne dans l'évolution et d'un rôle cellulaire fondamental. Celle-ci a une localisation cytoplasmique et catalyse la désamination des cytidines et des désoxycytidines libre en uridine ou désoxyuridine respectivement (Chung, Fromme, & Verdine, 2005). Celle-ci est aussi active sur certains analogues nucléotidiques utilisés comme traitements anticancéreux et antiviraux et pose donc un problème pharmacologique dans ces approches thérapeutique (Serdjebi, Milano, & Ciccolini, 2015). CDA est un homotétramère composée de 4 sous-unités identiques de 15 kDa (Figure 3) (Costanzi, Vincenzetti, Cristalli, & Vita, 2006). Chaque sous-unité lie un atome de zinc au niveau du site catalytique, essentiel à leur activité. Le fonctionnement de cette enzyme est assez complexe du fait que l'activité va être portée par le site actif d'une sous-unité complétée par certains résidus des trois autres sous-unités. CDA est exprimée et active dans le foie, la rate et dans une moindre mesure dans les poumons, les reins et dans les muqueuses du gros intestin. La concentration en CDA est très importante dans les neutrophiles matures et il a été montré que l'expression de CDA est un marqueur de la maturation des cellules myéloïdes dû à l'augmentation de l'expression d'ARNm de CDA durant cette maturation. Cette enzyme est retrouvée dans le liquide synovial d'individus atteints de polyarthrite rhumatoïde, relâché lors de la mort des neutrophiles (Mansson, Geborek, Saxne, & Bjornsson, 1990).

b. APOBEC/AID

La famille AID/APOBEC comprend 11 membres : APOBEC 1 (A1), A2, A3 -A, -B, -C, -DE, -F, -G, -H, A4 et AID (Figure 4A). Cette famille se distingue par deux caractéristiques : d'une part, elles arborent toutes un ou deux domaines catalytiques de coordination de zinc His/Cys-X-Glu-X₂₃₋₂₈-Pro-Cys-X₂₋₄-Cys (Harris & Liddament, 2004)(Figure 4B) conservé au cours de l'évolution à l'exception d'APOBEC4 dont le site catalytique diffère légèrement puisqu'il se termine par la séquence Pro-Cys-X6-Cys au lieu de Pro-Cys-X2-Cys (Rogozin, Basu, Jordan, Pavlov, & Koonin, 2005). D'autre part, leurs substrats sont les cytidines des acides nucléiques (Figure 4C).

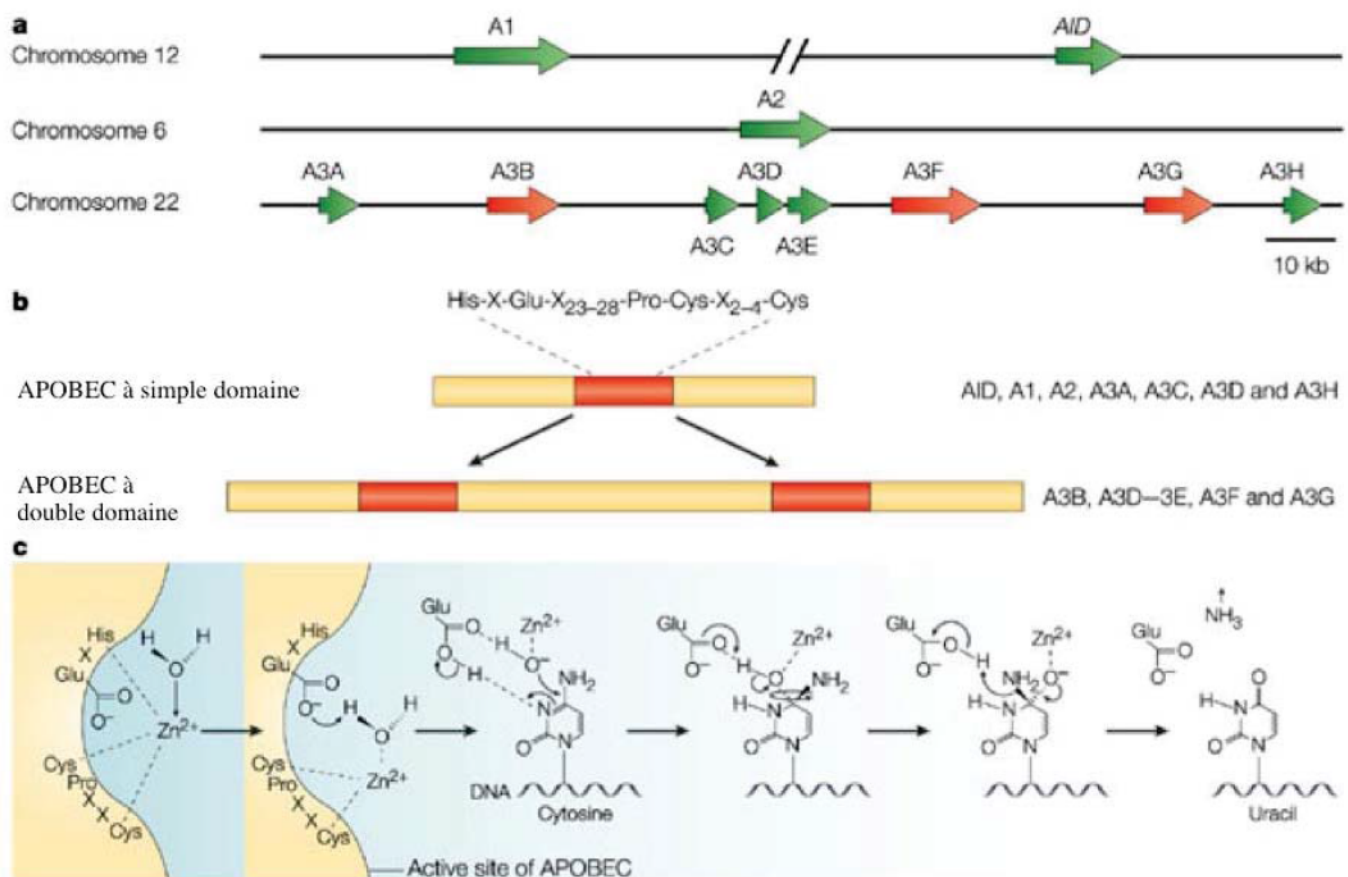


Figure 4 : Les différents membres de la famille APOBEC. A. Localisation des gènes codants pour les APOBEC/AID. Une flèche verte indique un simple domaine catalytique tandis que deux flèches rouges indiquent un double domaine. B. Organisation du ou des sites catalytique des APOBEC/AID. Le rectangle rouge indique un site catalytique dont la séquence consensus est indiquée en noir. C. Mécanisme de désamination par les APOBEC/AID. Tout d'abord, le site catalytique fixe un ion Zn^{2+} . L'action de cet ion et de l'acide glutamique va transformer une molécule d'eau en ion hydroxyde OH^- . L'acide glutamique va ensuite protoner le N_3 de la cytosine et déstabiliser la double liaison N_3-C_4 rendant C_4 vulnérable à l'attaque par l'ion OH^- . Ceci va entraîner ensuite le passage à un état intermédiaire tétraédrique qui va aboutir au clivage de la liaison C-N et à la formation d'une double liaison $C_4=O$ par la réaction du groupe amine avec le proton de la molécule d'eau. Enfin, le résidu uracile et une molécule d'ammoniaque NH_3 sont libérés du site actif. D'après Harris et al, 2004.

i. APOBEC 1

Le gène *APOBEC1*, localisé sur le chromosome 12, code la protéine APOBEC1 (A1) (Figure 4A) (Jarmuz et al., 2002). Cette protéine, chez l'Homme, est exclusivement exprimée dans le tractus intestinal. Ceci est expliqué par son rôle dans le métabolisme des lipides plasmatiques (Moris, Murray, & Cardinaud, 2014). Cette protéine est néanmoins retrouvée dans le foie chez la souris, le rat, le chien et le cheval (Greeve, Altkemper, Dieterich, Greten, & Windler, 1993). Le gène *APOBEC1* possède deux sites d'initiation de la transcription : l'un situé 955 bases en amont du codon d'initiation de la traduction et en aval d'une séquence Alu. L'autre site se situe dans l'intron 1, au sein d'une séquence Alu, délimitant le début de l'exon 2. Néanmoins, aucune différence n'est observée selon le promoteur utilisé car le codon d'initiation de la traduction se trouve au début de l'exon 2 (Fujino, Navaratnam, & Scott, 1998). En amont du premier promoteur se trouve un élément de réponse à p53 sur lequel la protéine p53 se fixe et induit la transcription d'*APOBEC1* (Ashur-Fabian et al., 2010).

Par ailleurs, un certain nombre de transcrits d'*APOBEC1* sont plus courts du fait d'une excision de l'exon 3. Néanmoins, ces ARNm dont la fonction est inconnue ne sont pas fonctionnels et sont éliminés par la cellule.

Jusqu'à récemment, la protéine A1 était seulement connue pour induire une mutation spécifique sur l'ARN messager de l'Apolipoprotéine B (ApoB) (B. Teng, Burant, & Davidson, 1993). On retrouve deux formes de la protéine ApoB: une forme longue, Apo-100 présente dans le foie, et une forme tronquée Apo-48, présente dans l'intestin. Ces deux protéines jouent un rôle dans le métabolisme lipidique. Les deux formes d'ApoB sont issues de deux ARN ne différant que d'une base : une cytidine dans l'ARNm de la protéine longue, mutée en uracile dans l'ARNm de la protéine tronquée. En effet, cette mutation en position 6666 conduit à la substitution d'un codon CAA (glutamine) en UAA (STOP) (Blanc, Kennedy, & Davidson, 2003; Powell et al., 1987). Ceci a conduit à la recherche d'un polymorphisme expliquant ce phénotype. Néanmoins, l'absence de mutation sur le gène *ApoB* a écarté l'hypothèse du polymorphisme. Ce résultat laissait donc supposer que c'était l'ARN qui était muté dans un processus post-transcriptionnel. C'est en recherchant l'enzyme à l'origine de cette mutation que fût découverte A1.

A1 se fixe avec peu d'efficacité sur l'ARNm. Elle a par conséquent besoin de cofacteurs qui vont former le pré-editosome 60S dans le cytoplasme (complexe non actif) qui

deviendra l'éditosome 27S dans le noyau (complexe actif). L'un de ces cofacteurs, l'A1 Complementation Factor (ACF) va dimériser et interagir à des dimères d'A1 afin de se fixer à l'ARNm de l'ApoB. ACF va reconnaître spécifiquement une séquence de 11 nucléotides appelée séquence d'ancrage en 3' de la cytidine en position 6666 (Smith, Bennett, Kizilyer, McDougall, & Prohaska, 2012). ACF assure à la fois la fixation à l'ARNm et la spécificité du substrat. Il a été montré que cette séquence d'ancrage est nécessaire et suffisante à l'activité d'A1 (Smith et al., 2012). L'*editing* a lieu dans le noyau et requiert donc l'import nucléaire du complexe. A1 possède une séquence de localisation nucléaire (NLS) en position Aminoterminal (N-ter) mais aussi un signal de rétention cytoplasmique en position carboxyterminal (C-Ter) (Chester et al., 2003; B. B. Teng et al., 1999; Y. Yang, Sowden, Yang, & Smith, 2001). ACF quant à elle possède à la fois un signal NLS et un signal d'export nucléaire (NES). La façon dont est régulé l'import nucléaire de ces deux protéines est encore mal connue mais il semblerait que la phosphorylation d'ACF joue un rôle important. En effet, seules les formes d'ACF phosphorylées sont retrouvées dans l'éditosome 27S. De plus, la coprecipitation d'ACF et d'A1 et l'activité d'*editing* de l'ARNm ApoB ne sont possibles que si ACF est phosphorylée. Dans le noyau, l'*editing* doit avoir lieu avant l'export de l'ARNm mais après l'épissage : ceci requiert donc l'import nucléaire d'A1 et d'ACF et que ceux-ci soient présents simultanément dans une fenêtre temporelle d'action très courte. L'activité d'A1 est donc très finement régulée (Smith et al., 2012). La surexpression d'A1 dans des modèles de souris transgéniques conduit à l'*editing* de nombreux ARNm, sur l'ADN génomique et à l'apparition de carcinomes hépatocellulaires associés à des dysplasies du foie, un taux important de lipides et à des mutations sur l'ARNm de la protéine tyrosine kinase TEC (Yamanaka et al., 1995). Ceci témoigne donc de l'importance de la régulation de l'expression et de l'activité d'A1.

A1 a donc un rôle clé dans la régulation du métabolisme lipidique et est impliquée dans des maladies telles que l'hyperlipidémie et l'athérosclérose.

Il a été montré qu'A1 pouvait aussi muter la cytidine en position 6802 sur le codon ACA et AUA, entraînant un changement d'une thréonine vers une isoleucine sur la protéine. Néanmoins, cette mutation n'a aucun effet sur la fonction de la protéine et son rôle reste inconnu (McDougall, Okany, & Smith, 2011). Une autre étude montre qu'A1 induit aussi une mutation sur l'ARNm de la neurofibromine 1 (NF1), un suppresseur de tumeur, en position

2914. La traduction de l'ARNm muté aboutit à la production d'une protéine tronquée retrouvée dans certaines tumeurs (Skuse, Cappione, Sowden, Metheny, & Smith, 1996).

Plus récemment, il a été montré chez l'Homme par RNAseq qu'A1 édite aussi la partie 3' UTR d'autres ARNm dans l'intestin grêle (Rosenberg, Hamilton, Mwangi, Dewell, & Papavasiliou, 2011). L'impact de ces mutations n'est en revanche pas encore connu.

ii. *APOBEC 2*

La protéine APOBEC2 (A2) a été découverte par recherche d'analogues avec A1 (Liao et al., 1999). La protéine A2 est codée par le gène *APOBEC2* situé sur le chromosome 6 (Figure 4A) (Jarmuz et al., 2002). A2 est exprimée dans le tissu cardiaque et le muscle squelettique (Liao et al., 1999). Néanmoins, le rôle d'A2 demeure inconnu car aucune activité cytidine désaminase n'a pu être mise en évidence dans le modèle bactérien ou *in vitro* (Liao et al., 1999). Il s'agit de la première APOBEC dont toute la structure cristallographique a été déterminée (Prochnow, Bransteitter, Klein, Goodman, & Chen, 2007). Celle-ci a permis l'étude de la structure des autres membres par homologie. Il a été montré que des souris déficientes pour A2 présentaient une baisse de corpulence et des myopathies ce qui suggère un rôle essentiel de cette protéine dans le développement musculaire (Y. Sato et al., 2010). Par ailleurs, A2 est surexprimée dans les hépatocytes sous l'action des cytokines pro-inflammatoires TNF α et IL1 β (Vonica, Rosa, Arduini, & Brivanlou, 2011). Ceci serait dû au fait qu'A2 possède un élément de réponse à NF κ B fonctionnel situé en 5'UTR. Une étude plus récente montre qu'une surexpression d'A2 chez la souris entraîne l'apparition de tumeurs du foie et des poumons. Ceci serait dû à une activité d'*editing* d'A2 spécifiquement sur des gènes suppresseurs de tumeur (Okuyama et al., 2012). Ces résultats semblent donc indiquer qu'A2 pourrait jouer un rôle important dans le développement musculaire et dans l'apparition de carcinomes hépatocellulaires dans des conditions d'inflammation chronique.

iii. *APOBEC 4*

La protéine APOBEC 4 (A4) est codée par le gène *APOBEC4* situé sur le chromosome 1 et est exprimée dans les testicules (Figure 4A) (Rogozin et al., 2005). Sa localisation est nucléaire. La protéine A4 diffère des autres APOBEC par l'originalité de son site catalytique hypothétique : en effet, son domaine de coordination en doigt de zinc se

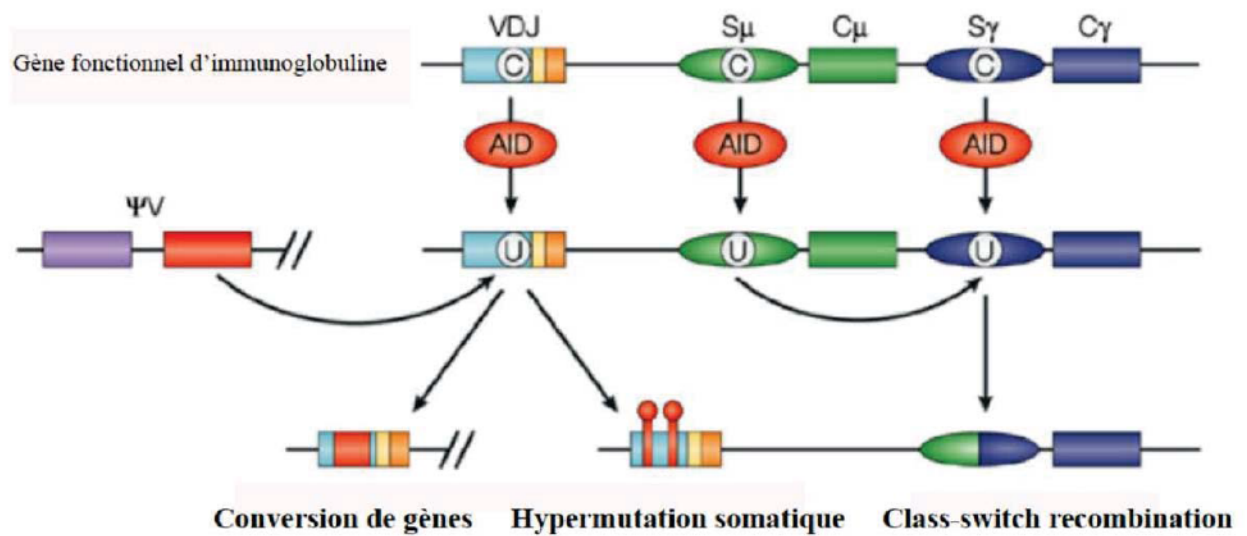


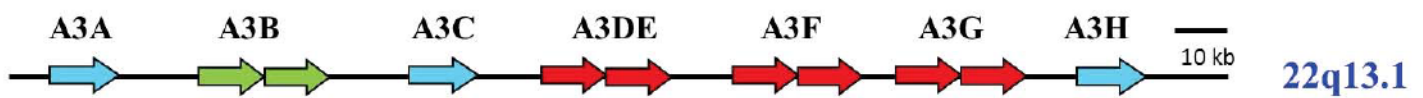
Figure 5 : Mécanismes de diversification des gènes d'immunoglobulines par AID. Sont ici représentés de façon schématique les trois mécanismes permettant la diversification du répertoire des immunoglobulines par AID : la conversion de gènes, les hypermutations somatiques et la recombinaison de classe (Class- switch recombination). D'après Harris et al, 2004.

termine par la séquence Pro-Cys-X₆-Cys au lieu de Pro-Cys-X₂₋₄-Cys (Rogozin et al., 2005). Aucune activité cytidine désaminase n'a pu lui être attribuée chez l'Homme. Des expériences menées dans le modèle bactérien ou de levure n'ont pas permis d'attribuer un rôle mutagène à A4. Le rôle d'A4 demeure donc inconnu. On pense aujourd'hui que son activité pourrait nécessiter la présence d'un cofacteur, ou alors que sa fonction n'est pas liée à une activité cytidine désaminase.

iv. AID

La protéine Activation Induced Cytidine Deaminase (AID ou AICDA) a été mise en évidence par hybridation soustractive en comparant les transcrits obtenus sur des cellules B murines activées et non activées (Muramatsu et al., 1999). AID est une protéine de 198 acides aminés. Elle est codée par le gène *AICDA* situé sur le chromosome 12 (Figure 4A) (Jarmuz et al., 2002). Cette protéine est très conservée au sein des vertébrés (Harris & Liddament, 2004). Bien qu'AID remplisse son rôle physiologique dans le noyau, elle est retrouvée dans le cytoplasme du fait de l'interaction entre l'exportine CRM1 et son signal d'export nucléaire (NES), constitué d'une région d'une dizaine d'acides aminés riches en leucine à l'extrémité C-terminal de la protéine. Son import dans le noyau se fait de façon active par interaction entre l'importine alpha et le signal de localisation nucléaire (NLS) d'AID situé dans sa région N-Terminal. AID est principalement exprimée dans les cellules lymphocytaires B (Patenaude & Di Noia, 2010). Elle joue un rôle dans la maturation du répertoire des anticorps par deux mécanismes. D'une part, elle participe à la commutation de classe des immunoglobulines (commutation isotypique). D'autre part, elle est directement impliquée dans la diversification des immunoglobulines par l'induction d'hypermutations somatiques sur les gènes codant les régions variables de celles-ci (Muramatsu et al., 2000)(Figure 5). Ce mécanisme contribue à l'augmentation de l'affinité des anticorps pour un antigène donné. Cette activité est régulée par la phosphorylation de la thréonine 27 par la Protein Kinase A (PKA). AID a donc un rôle clé dans la mise en place de l'immunité adaptative. Des mutations sur le gène AID ont été observées chez les patients souffrant du syndrome hyper IgM : l'expression d'une protéine AID non fonctionnelle aboutit à la production de grandes quantités d'IgM et à une absence d'expression des autres isotypes (Minegishi et al., 2000; Revy et al., 2000). Ceci est aussi observé chez les souris AID *knock-out*. De par son homologie de séquence avec A1, on a longtemps considéré que le substrat d'AID était l'ARN : il a été montré qu'AID a pour substrat l'ADN simple brin (Harris, Sale, Petersen-Mahrt, & Neuberger, 2002; Petersen-

A.



B.

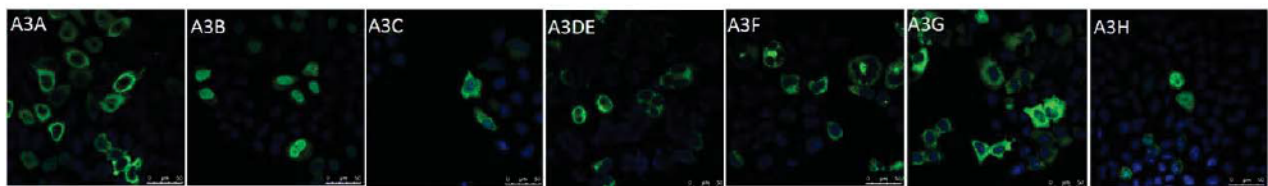


Figure 6 : Organisation et localisation cellulaire des APOBEC3. A. Représentation schématique du locus APOBEC3 sur le chromosome 22. Une flèche indique une A3 à simple domaine tandis que deux flèches indiquent une A3 à double domaine. La localisation cellulaire est indiquée par la couleur : bleue pour une localisation nucléo cytoplasmique, verte pour une localisation nucléaire et rouge pour une localisation cytoplasmique. B. Localisation cellulaire des A3 par microscopie à fluorescence. Les A3 couplés à un tag V5 sont transfectés dans les cellules HeLa et observés par microscopie confocale à fluorescence.

Mahrt & Neuberger, 2003) mais aussi, dans une moindre mesure, l'ARN simple brin (Liang et al., 2013). Il est par conséquent nécessaire que la cible soit accessible et sous forme simple brin. C'est notamment le cas au moment de la transcription des gènes. Il a été montré qu'une interaction entre AID, l'ARN polymérase II (RNAPol II) et la protéine RNA polymérase-associated factor 1 (PAF1) contribuait à l'accès d'AID au locus des immunoglobulines (Ig). C'est ainsi qu'AID va désaminer les cytidines en uracile sur la région variable (V) des Ig. Ces mutations ont lieu sur les motifs WRCY (W= adénine ou thymine, R= purine et Y= pyrimidine). AID peut aussi induire des cassures de l'ADN double brin (DSBs) dans la région switch (S) du gène codant les immunoglobulines: c'est ce phénomène qui est à l'origine de la commutation isotypique. Ces deux phénomènes font appel ensuite aux systèmes de réparation de l'ADN, notamment le système de réparation des mésappariements (MMR) que nous détaillerons plus tard.

Par ailleurs, il a été montré que AID est aussi exprimée dans les ovocytes, les cellules souches embryonnaires et dans les tissus de seins sensibles aux œstrogènes (Moris et al., 2014). Le rôle d'AID dans ces cellules n'est pas encore connu. Néanmoins, des études chez le zebrafish suggèrent qu'AID est impliquée dans la reprogrammation des cellules germinales durant les phases précoces du développement (Rai et al., 2008). De plus, l'expression constitutive chez des souris transgénique pour AID entraîne une durée de vie plus courte des souris et l'apparition de tumeurs : cet aspect sera traité plus en détail dans une prochaine partie.

D'autre part, il a été montré qu'AID est active sur les résidus 5-méthylcytosine (5-meC) ce qui suggère un rôle dans la reprogrammation épigénétique des cellules.

AID a donc un rôle crucial dans la production d'anticorps mais aussi dans la régulation épigénétique. De par son activité nucléaire, on suppose qu'AID peut être impliquée dans l'apparition de cancers et qu'elle doit donc être particulièrement bien régulée.

v. *APOBEC3*

Les A3s forment une famille composée de 7 membres chez l'Homme (A3A, A3B, A3C, A3DE, A3F, A3G et A3H). Les gènes codant les A3s forment un locus situé sur le chromosome 22 (Jarmuz et al., 2002)(Figure 6A). Ces cytidines désaminases sont toutes actives, à l'exception d'A3DE et ont pour substrat l'ADN simple brin. Par ailleurs, il a été

montré que A3A et A3G peuvent se fixer à l'ARN mais seule A3A est capable d'induire des mutations sur l'ARN, dans les macrophages M1 (Sharma et al., 2015).

α . APOBEC3A

A3A est une protéine de 181 acides aminés et d'un poids moléculaire d'environ 20kDa. Avant d'être identifiée en tant que cytidine désaminase, APOBEC3A (A3A) était connue sous le nom de phorbolin-1 car elle est surexprimée par le phorbol12-myristate 13-acetate (PMA) ou le probol 12-13 dibutyrate (PDBu)(Madsen et al., 1999). La protéine A3A n'est formée que d'un domaine catalytique : ceci lui permet de diffuser passivement par les pores nucléaires. A3A a par conséquent une localisation nucléo cytoplasmique (Figure 6B). En plus de son activité mutatrice, A3A peut aussi induire des cassures de l'ADN double brin (DSBs) (Landry, Narvaiza, Linfesty, & Weitzman, 2011).

L'ARNm d'A3A est très peu stable en raison de la présence d'une séquence Alu dans sa partie 3' non traduite (3'-UTR) (Caval, Suspene, Shapira, Vartanian, & Wain-Hobson, 2014). De plus, cet ARNm comporte deux codons d'initiation situés dans un bon contexte de Kozak et codent deux isoformes d'A3A : P1 et P2. Ces deux isoformes ne diffèrent que de 12 acides aminés : ceci s'explique par la présence de deux codons d'initiation situés dans un bon contexte de Kozak. La forme P2 est issue d'une initiation interne de la traduction et est donc plus courte que P1 (Mussil et al., 2013).

La protéine A3A est formée de cinq feuillets beta et de six hélices alpha. A3A peut former des homéomères par sa partie N-terminal : ceci permet la mise en commun des résidus chargés positivement dans la poche de fixation à l'ADN et d'améliorer l'affinité d'A3A pour son substrat (Bohn et al., 2015). Par ailleurs, la fixation de l'ADN va au-delà du site catalytique. De plus, A3A se fixe d'autant mieux à l'ADN si celui-ci est long, suggérant que d'autres résidus peuvent être impliqués dans la fixation à l'ADN. La souplesse de la boucle 3, reliant le feuillet beta 2 à l'hélice alpha 2, va entraîner un changement de conformation de la protéine lors de sa fixation à son substrat. Cette souplesse de la boucle 3 est importante pour l'activité d'A3A mais le détail mécanistique de ce changement de conformation reste encore inconnu.

L'expression d'A3A est très régulée : celle-ci est exprimée dans les lymphocytes du sang périphérique, les kératinocytes et la rate (Refsland et al., 2010). Par ailleurs, il a été montré que l'IFN α induisait l'expression d'A3A dans les monocytes et macrophages, les cellules dendritiques et les lymphocytes T CD4⁺ (Aynaud et al., 2012; Refsland et al., 2010). La (PMA) induit aussi une surexpression d'A3A dans les monocytes et les kératinocytes (ref).

L'activité d'A3A est aussi régulée par l'interaction avec la protéine Tribbles homolog 3 (TRIB3). TRIB3 est une pseudokinase, proche des serine/thréonine kinases tel que PIM1. TRIB3 est impliquée dans de nombreux mécanismes cellulaires : il a été montré que celle-ci inhibait la phosphorylation de la protéine AKT et son activation par l'insuline pouvant aboutir au diabète et à des maladies cardiovasculaires. Elle est aussi impliquée dans la régulation du cycle cellulaire, notamment aux points de contrôle G1/S et G2/M (Sakai et al., 2010). Nous avons montré dans le laboratoire que TRIB3 régule l'activité d'A3A en entraînant la dégradation de celle-ci. Cette dégradation est spécifique à A3A bien que TRIB3 puisse aussi interagir avec A3C. Ceci se traduit *in vitro* par une diminution de l'activité mutagenétique d'A3A et des DSBs (Aynaud et al., 2012).

Plus récemment, il a été montré qu'A3A induit des mutations C->U sur l'ARNm dans des monocytes issus de PBMCs ou dans les macrophages M1 stimulés par l'interféron de type I ou en hypoxie. En effet, dans ces conditions, la fréquence de mutations C->U sur les ARNm est plus importante, notamment sur l'ARNm de la succinate déshydrogénase B (SDHB). Par comparaison avec d'autres lignées de macrophage, les auteurs montrent que l'*editing* de l'ARNm de SDHB est corrélé à une forte expression d'A3A. De plus la surexpression d'A3A dans la lignée 293T induit une augmentation de l'*editing* des ARNm tandis que le silencing de l'expression d'A3A dans la lignée M1 réduit l'*editing* des ARNm. Ceci montre par ailleurs qu'A3A semble préférentiellement éditer les cytidines situées sur les boucles des structures tige-boucle des ARNm (Sharma et al., 2015). Ceci a également été montré sur l'ARNm de WT1, un suppresseur de tumeur, dans les cellules sanguines mononuclées de cordon ombilical. La surexpression d'A3A dans ces cellules est corrélée à une fréquence plus élevée en mutations sur l'ARNm WT1 tandis que le silencing d'A3A s'accompagne d'une diminution des mutations (Niavarani et al., 2015).

β. APOBEC3B

A3B est une protéine d'environ 48 kDa. Il s'agit d'une A3 possédant 2 domaines catalytiques : l'un en position aminoterminal (N-ter) et l'autre en position carboxyterminale (C-ter) (Jarmuz et al., 2002). Néanmoins, seul le domaine C-ter est actif. A3B a une localisation exclusivement nucléaire (Figure 6B). Ceci s'explique par la présence d'un NES en N-ter. A l'instar d'A3A, A3B peut aussi induire des mutations sur l'ADN nucléaire, sur les résidus 5-MeC mais A3B semble néanmoins moins active qu'A3A. En revanche, A3B n'induit pas de DSBs (Caval, Suspene, Shapira, et al., 2014). Les déterminants de cette atténuation de l'activité d'A3B par rapport à A3A font l'objet d'une étude présentée dans la partie résultats. A3B est retrouvée en faible ou moyenne quantité dans le foie, les ovaires, les poumons, la rate, le colon et le cœur (Refsland et al., 2010). A3B est aussi retrouvée surexprimée dans certains cancers du sein. Ceci est associé à un taux de mutations C->T plus important. Peu étudiée jusqu'à récemment, A3B est maintenant considérée comme une A3 impliquée dans de nombreux cancers (Burns, Lackey, et al., 2013; Burns, Temiz, & Harris, 2013). Néanmoins, une délétion de 29,5 kb entre les gènes *APOBEC3A* et *APOBEC3B*, particulièrement présente dans les populations Asiatiques et d'Océanie, entraîne la production d'un ARNm d'A3A chimère dont la séquence 3'UTR correspond à celle d'A3B (Caval, Suspene, Shapira, et al., 2014; Kidd, Newman, Tuzun, Kaul, & Eichler, 2007). Il en résulte un ARNm beaucoup plus stable. De façon intéressante, cette délétion semble être associée à un certain nombre de cancers. Nous développerons le lien entre A3B et cancer dans une prochaine partie.

γ. APOBEC3C

A3C est une A3 monodomaine de poids moléculaire proche d'A3A. Elle a une localisation nucléo-cytoplasmique mais n'est pas active sur l'ADN nucléaire (Figure 6B). Elle reste néanmoins active sur l'ADN mitochondrial et les plasmides transfectés (Suspène, Aynaud, Guétard, et al., 2011). A3C est beaucoup plus ubiquitaire puisqu'elle est retrouvée dans la vessie, le cervix, le cœur, le poumon, les ovaires, la rate et les leucocytes du sang périphérique : il s'agit d'une des A3 les plus abondantes chez l'Homme (Refsland et al., 2010). A3C est composée, tout comme A3A de six hélices alpha et de 5 feuillets beta. De plus, elle forme des homodimères par l'intermédiaire de ses résidus F55 et W74 : la mutation

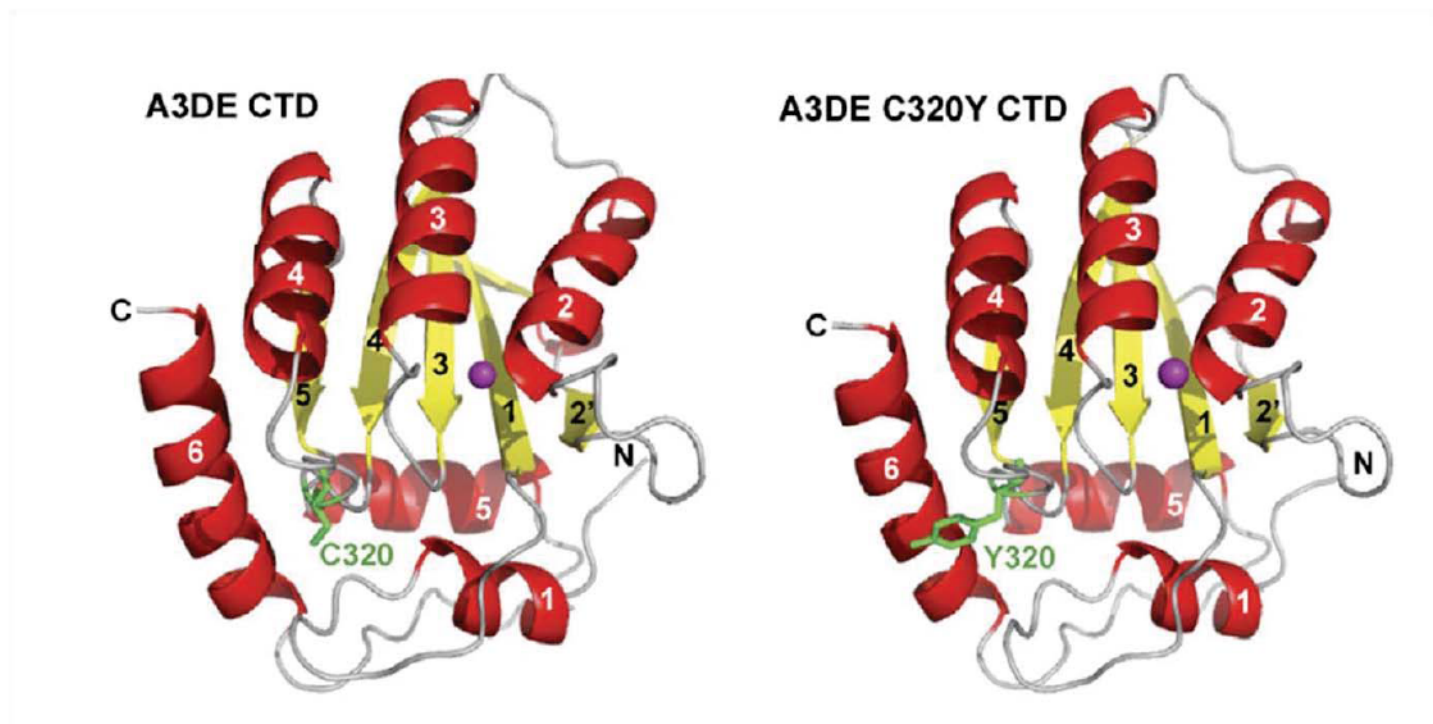


Figure 7 Représentation de la structure prédite de hA3DE classique et hA3DE C320Y. Les structures rouges représentent les hélices alpha tandis que les jaunes les feuillets beta. L'acide aminé en position 320 est représenté en vert et l'atome de zinc en violet. Les structures sont prédites à l'aide YASARA et PYMOL. D'après Dang et al, 2011.

de ses positions abolit la restriction du Virus de l'Immunodéficience Humaine (VIH) démontrant ainsi que cette interaction est nécessaire à l'activité d'A3C (Stauch et al., 2009). Peu de choses sont connues en revanche sur la régulation d'A3C. On sait néanmoins qu'A3C interagit avec TRIB3 sans que cela ait d'effet sur la stabilité de la protéine (Aynaud et al., 2012).

δ. APOBEC3DE

A3DE est une A3 à double domaine d'environ 50 kDa. Celle-ci a une localisation strictement cytoplasmique (Figure 6B). A3DE possède la particularité d'être peu ou pas active, entraînant une très faible activité de restriction antivirale de sa part. Cette absence d'activité est due à la présence d'une cystéine en position 320. La substitution de cette cystéine par une tyrosine suffit à rétablir une activité cytidine désaminase. On suppose que cette cystéine gêne la fixation d'A3DE à l'ADN par encombrement stérique (Figure 7) (Dang et al., 2011). D'après les bases de données, seuls 0,1% de la population mondiale possèdent cet allèle fonctionnel d'A3DE. Néanmoins, la forme active est retrouvée chez un certain nombre de primates ce qui indique une évolution récente vers l'inactivation chez l'Homme. Nous détaillerons cet aspect plus en détail dans une partie dédiée.

A3DE est exprimée dans la vessie, le cervix, le foie, le poumon, la rate, le thymus et les lymphocytes du sang périphérique (Refsland et al., 2010). A3DE est aussi surexprimée dans des tissus cirrhotiques infectés par le Virus de l'Hépatite B (VHB), le Virus de l'Hépatite C (VHC) ou co-infectés par le VHB et le VHC mais A3DE étant dépourvue d'activité, son rôle physiologique demeure inconnu (Figure 8) (Vartanian et al., 2010). Nous avons néanmoins mené des travaux qui montrent qu'A3DE peut jouer un rôle de modulateur d'A3F et d'A3G. Ceci est détaillé dans la partie résultats.

ε. APOBEC3F

A3F est une A3 très proche d'A3DE et de taille similaire. Tout comme A3DE, elle est composée de deux domaines et a une localisation cytoplasmique (Figure 6B). De façon intéressante, A3F est aussi retrouvée dans les P-bodies, structures de stockage et de dégradation d'ARNm non traduits. Néanmoins, le rôle d'A3F dans ces structures n'est pas connu. A3F peut induire des mutations sur l'ADN mitochondrial. A3F a surtout été étudiée en

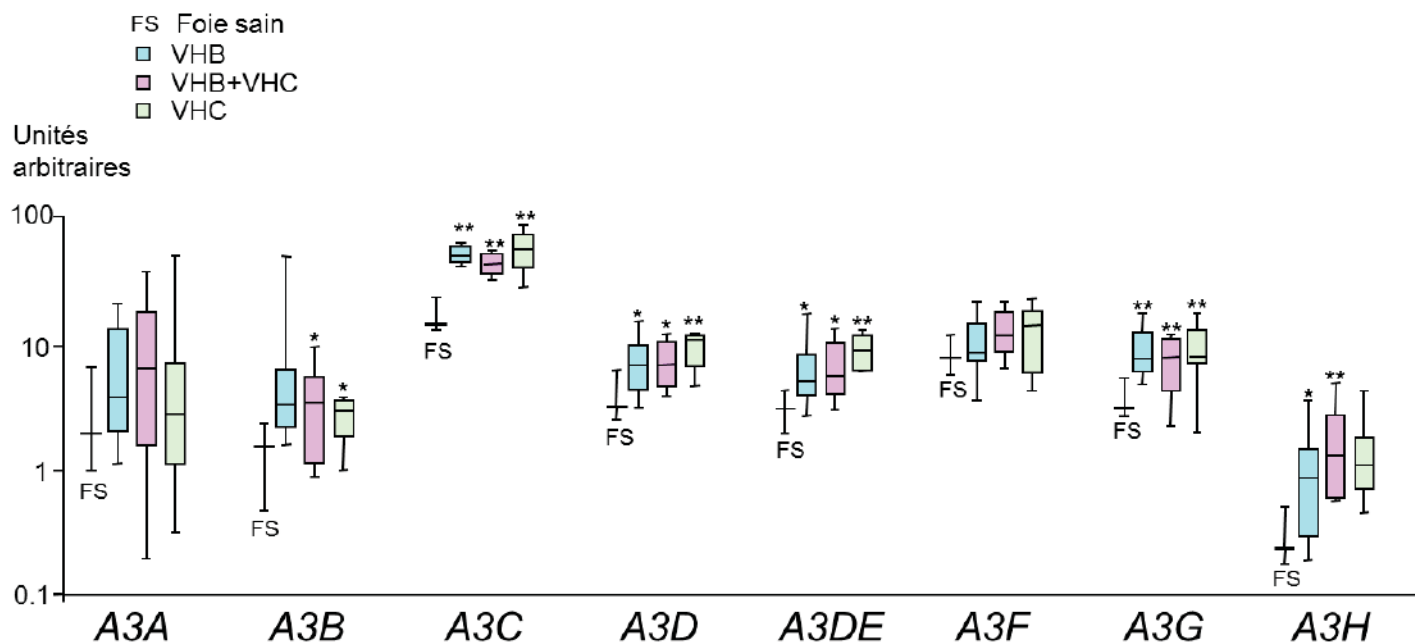


Figure 8 : Transcriptome des APOBEC3 sur des tissus cirrhotiques de patients sains et patients infectés par VHB, le VHC ou co infectés VHB et VHC. Les données sont normalisées sur la base de gènes de références dont l'expression est invariable (TRIM44, HMBS et LMF2). Les astérisques indiquent une différence significative : **0.01<p<0.001 ; *0.05<p<0.01. D'après Vartanian et al. 2010

tant que facteur de restriction contre le Virus de l'Immunodéficience Humaine (VIH). Nous détaillerons l'activité antivirale d'A3F dans une prochaine partie.

A3F est exprimée dans la vessie, le cervix, le foie, le poumon, la rate, le thymus et les lymphocytes du sang périphérique (Refsland et al., 2010). Certaines études montrent qu'A3F est surexprimée par les cellules dendritiques immatures stimulées à l'interféron alpha. (Mohanram, Skold, Bachle, Pathak, & Spetz, 2013)

ζ. APOBEC3G

A3G est une A3 à double domaine et de localisation cytoplasmique (Figure 6B). Son poids moléculaire est semblable à celui d'A3F. A3G est l'une des A3 les plus étudiées, du fait qu'elle ait été découverte en tant que facteur de restriction contre le VIH-1. C'est dans ce contexte qu'elle a été identifiée pour la première fois en 2002 sous le nom CEM15 : il était connu à l'époque que le VIH-1 déficient pour la protéine Vif ne peut pas se répliquer dans les cellules dites non permissives. Par la technique d'hybridation soustractive, il a été montré que la protéine conférant la non permissivité au VIH-1 delta Vif est CEM15 et que ce facteur est contrecarré par la protéine Vif (Sheehy, Gaddis, Choi, & Malim, 2002). A3G est aussi active sur l'ADN mitochondrial {Suspène, 2011 #1053}

A3G est exprimée dans la vessie, le cervix, le foie, le poumon, la rate, le thymus et les lymphocytes du sang périphérique (Refsland et al., 2010). Certaines études montrent qu'A3G est surexprimée par les cellules dendritiques immatures, les macrophages et les cellules hépatiques stimulées à l'interféron alpha (Bonvin et al., 2006; Koning et al., 2009; Tanaka et al., 2006; F. X. Wang, Huang, Zhang, & Ma, 2008). Ceci s'explique par la présence de deux séquences Interferon Regulatory Factor Element (IRF-E) putatives dans la région -128/+7 du promoteur d'A3G. Il a été montré, en plaçant cette séquence en amont d'un gène rapporteur (la luciférase en l'occurrence) que l'interféron alpha permet une surexpression de la luciférase d'environ 4 fois comparé au contrôle. Ces séquences semblent donc être responsables de la surexpression d'A3G lors de stimulation par l'interféron (Tanaka et al., 2006). Par ailleurs, l'étude du promoteur d'A3G révèle la présence de séquences de fixation des facteurs de transcription Sp1 et Sp3 en -87/-78. Ceux-ci semblent impliqués dans l'expression basale d'A3G. En effet, l'emploi d'un siRNA contre Sp1 ou Sp3 aboutit à une diminution de l'expression d'A3G y compris dans les cellules ne surexprimant pas A3G lorsqu'elles sont stimulées par l'interféron. A3G est donc régulé par deux systèmes : l'un lié à l'expression



Figure 9 : Représentation schématique des A3 en fonction de leur type de domaine catalytique Z1, Z2 ou Z3. D'après LaRue et al, 2009.

basale d'A3G (Sp1/Sp3) et l'autre spécifique des situations d'inflammation dans les cellules sensibles à l'interféron pour A3G (séquences IRF-E) (Muckenfuss et al., 2007).

η. APOBEC3H

A3H, dernier membre de la famille A3, ne possède qu'un seul domaine catalytique. Celle-ci existe sous forme d'au moins sept haplotypes (I à VII) qui sont définis par des polymorphismes retrouvés dans les exons 2, 3 et 4. Sa localisation, principalement nucléocytoplasmique (Figure 6B), varie selon l'haplotype considéré : en effet, l'haplotype I a par exemple une localisation plutôt nucléaire tandis que l'haplotype II est plutôt cytoplasmique. Il a été montré que l'entrée dans le noyau se fait de façon passive tandis que la rétention dans le cytoplasme fait intervenir un mécanisme actif.

A3H a été étudiée dans le contexte de l'infection au VIH et la restriction des rétroéléments transposables. La restriction du VIH-1 se fait aussi bien par son activité cytidine désaminase qu'indépendamment de son activité catalytique. A3H est aussi capable de désaminer les cytidines de l'ADN mitochondrial. L'étude de la structure d'A3H, notamment de l'haplotype II, révèle que le mécanisme d'action d'A3H est plus proche des A3 à double domaine que des autres A3 à simple domaine. Ces différents haplotypes diffèrent dans la stabilité de la protéine, ce qui permet de classer ces haplotypes en deux catégories : les haplotypes II, V et VII qui sont qualifiés de stables et les haplotypes I, III, IV et VI qui sont qualifiés d'instables. Ces formes sont impliquées dans la résistance ou la sensibilité à certains variant du VIH-1 : les haplotypes codant des formes instables d'A3H semblent conférer une plus grande sensibilité à l'infection par le VIH-1 que les formes stables. De façon intéressante, les haplotypes stables sont plus fréquents dans les populations africaines tandis que les haplotypes instables sont plus fréquents dans les populations asiatiques (OhAinle, Kerns, Li, Malik, & Emerman, 2008; OhAinle, Kerns, Malik, & Emerman, 2006).

A3H est exprimée dans les PBMC, le colon, le poumon, la rate, le thymus et le sein (ref) mais aussi dans les CD4⁺ stimulés (Refsland et al., 2010).

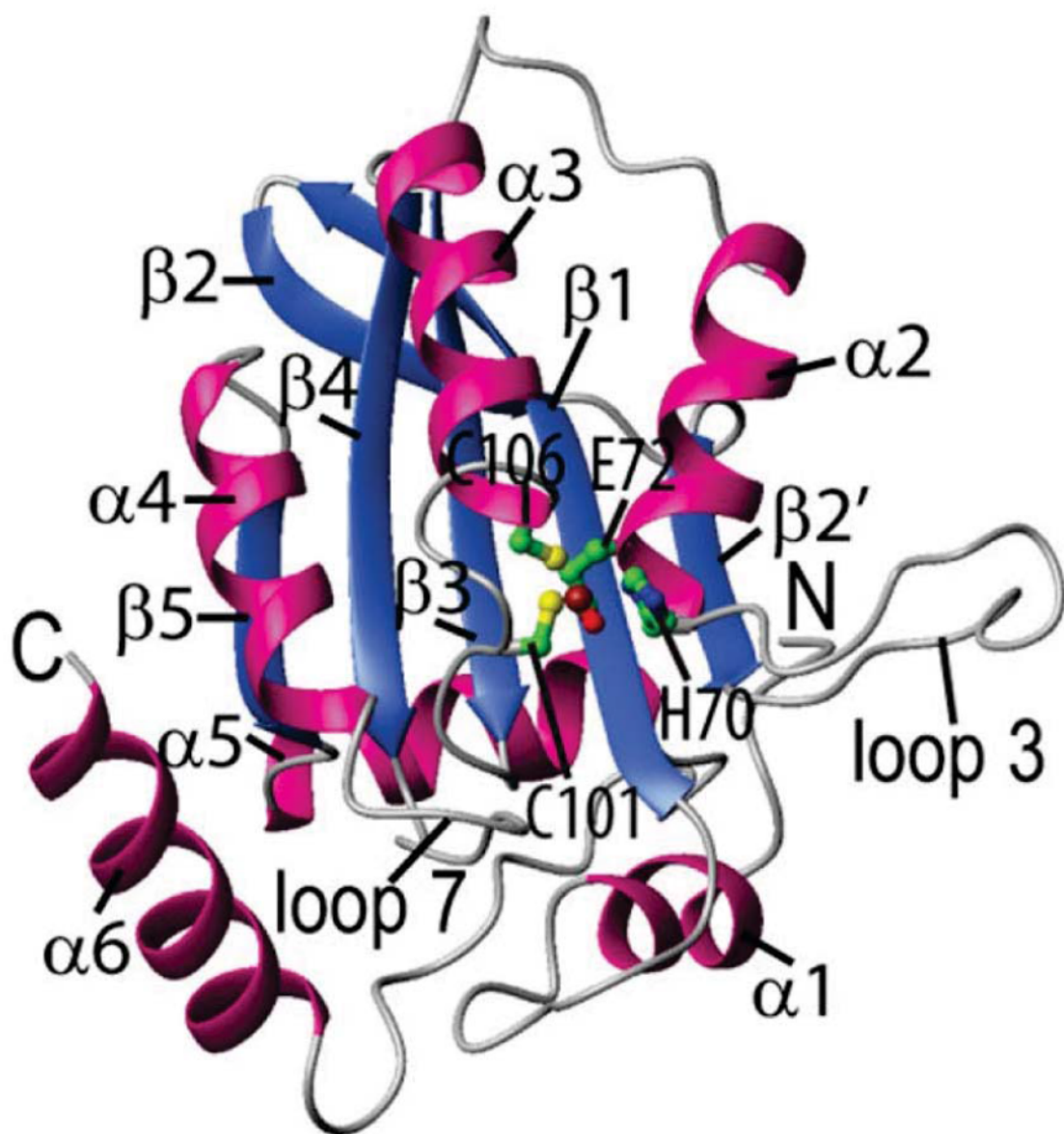


Figure 10 : Modèle de la structure d'A3A obtenue par NMR

Ce modèle est le plus stable des 30 conformères obtenus par NMR. Les hélices α sont colorées en rose tandis que les feuillets β sont colorés en bleu. Les extrémités, éléments structuraux (hélices α , boucles et feuillets β) et résidus du site actif (H70, E72, C101 et C106) sont indiqués en noir. L'ion zinc, les atomes de carbone, d'azote, d'oxygène et de soufre sont représentés sous forme de « ball and stick » et colorés en marron, vert, bleu, rouge et jaune respectivement. D'après Byeon et al, 2013.

vi. Structure des A3

La grande homologie de séquence entre les A3 de primates permet de classer les A3 selon le ou les domaines qui les composent. On classe le ou les domaines de ces A3 en 3 catégories : Z1, Z2 et Z3. Les domaines de type Z1 forment la protéine A3A, la partie carboxy-terminale d'A3B et la partie carboxy-terminale d'A3G. Z2 constitue la partie amino-terminale d'A3B, A3C, les deux domaines d'A3DE et d'A3F et enfin le domaine amino-terminal d'A3G. Enfin, A3H est constituée d'un domaine Z3 (Figure 9) (Larue, Andresdottir, et al., 2008).

Ces homologies de séquence sont impliquées dans la formation d'homodimères et d'oligomères, comme c'est le cas pour A3G et A3A. Des mesures par luminescence montrent qu'A3B, A3DE, A3F et A3H pourraient aussi former des dimères ou des oligomères dans des complexes de haut poids moléculaire (Li et al., 2014). Il existe aussi des hétérodimères comme c'est le cas pour A3F et 3G (Wiegand, Doehle, Bogerd, & Cullen, 2004). Ceci laisse supposer que les A3 partageant un ou deux domaines homologues peuvent former des hétérodimères. Néanmoins, la fonction de ces interactions reste inconnue. Sur la base des homologies de séquences, nous avons vérifié si A3DE interagit avec d'autres A3 et quelle pouvait être la fonction d'une telle interaction. Nous développerons cet aspect dans la partie résultats.

La structure tridimensionnelle de chaque APOBEC n'est pas encore connue. Seules les structures cristallographiques d'A2 et d'A3C sont connues. Des structures partielles d'A3G et d'A3F ont néanmoins été obtenues. Enfin, la structure d'A3A a été déterminée par résonance magnétique nucléaire (NMR) (Figure 10) (Byeon et al., 2013). Ces données ont permis de prédire des modèles pour d'autres A3. Globalement, tous les monodomaines sont constitués de six hélices alpha organisées autour de 5 feuillets beta. Le motif de coordination en doigt de zinc se situe au niveau de $\alpha 2$ - $\beta 3$ - $\alpha 3$ tandis que le site catalytique se situe au niveau des hélices $\alpha 5$ - $\alpha 6$ (Figure 10).

vii. Evolution des cytidines désaminases APOBEC/AID

Les deux premiers représentants de la famille APOBEC/AID sont A2 et AID. Celles-ci sont apparues il y a 425 millions d'années environ chez les poissons tels que le zebrafish

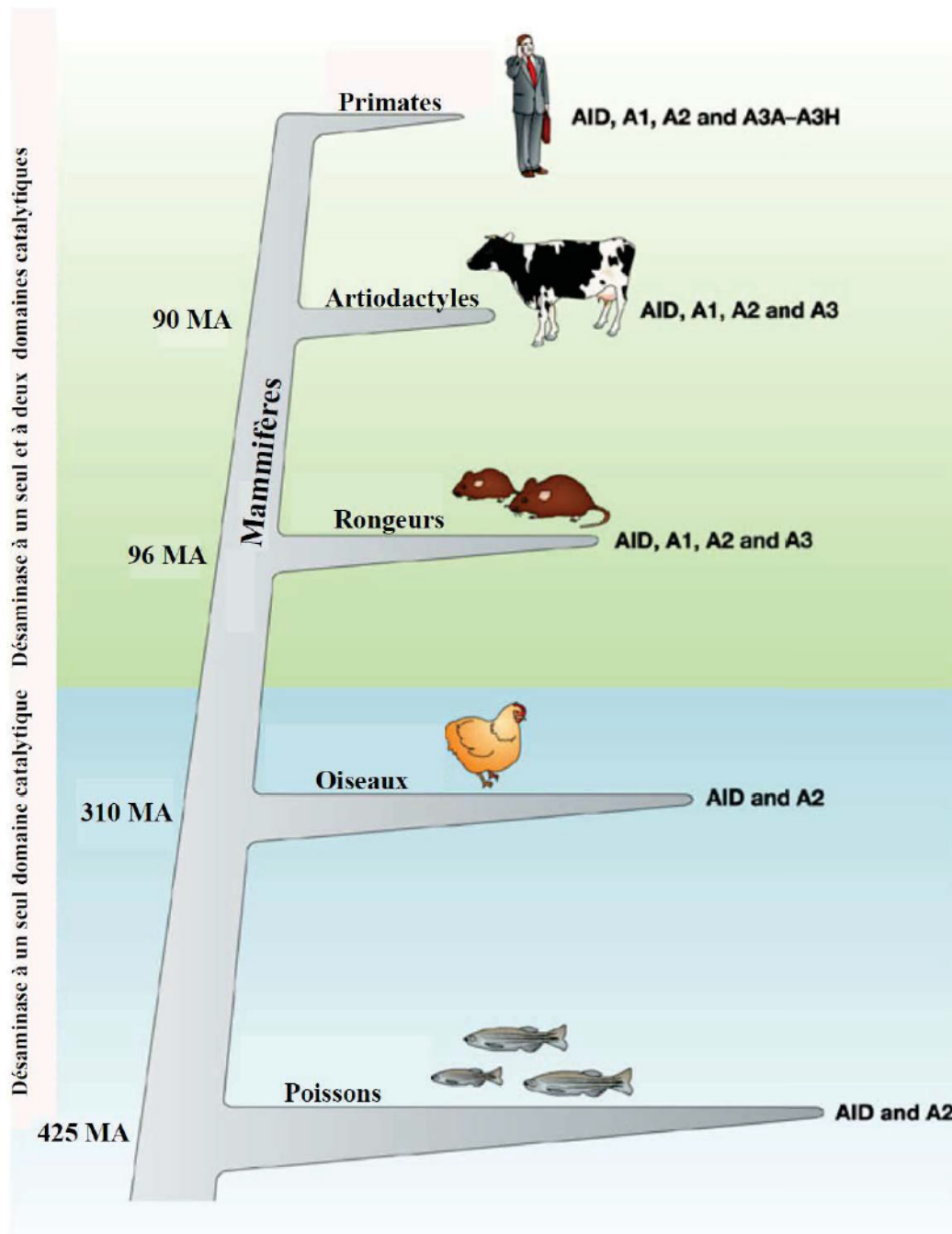


Figure 11 : Evolution des APOBEC/AID chez les vertébrés. Représentation de la phylogénique des grands groupes de vertébrés et des APOBEC/AID. MA : Millions d'années. D'après Harris et al, 2004.

(Figure 11). Ces deux protéines sont fortement conservées au cours de l'évolution puisque l'AID humaine et l'AID de poisson zèbre partagent 70 à 74% de similitudes tandis que l'A2 humaine et l'A2 de poisson zèbre partagent 68 à 70% de similitudes (Harris & Liddament, 2004).

Les rongeurs sont le premier groupe dans l'évolution à posséder, en plus d'AID et d'A2, une A1 et une A3 (Figure 11). Cette A3 est aussi la première des APOBEC/AID à être composée de deux domaines catalytiques. Cette émergence date d'environ 96 millions d'années (ref). Les rongeurs ne possèdent qu'une seule A3, située sur le chromosome 15. Les rongeurs possèdent donc quatre cytidines désaminases : AID, A1, A2 et A3 (composée d'un domaine Z2 et d'un domaine Z3), de même que les artiodactyles. Ce nombre augmente fortement pour atteindre 11 cytidines désaminases chez les primates par l'apparition du gène *A4* et des différents gènes *A3* (*A3A*, *A3B*, *A3DE*, *A3F*, *A3G* et *A3H*) (Figure 11). Ces différentes A3 seraient apparues au cours de l'évolution par duplication de gènes, ce qui confère aux A3 une forte homologie de séquences (Harris & Liddament, 2004). On observe par exemple que les gènes *A3A*, composé d'un domaine Z1 chez les mammifères, sont orthologues : leur séquence est proche, ils possèdent tous l'activité cytidine désaminase, la capacité de désaminer les 5 MeC et d'induire des cassures de l'ADN double brin (Caval, Suspene, Vartanian, & Wain-Hobson, 2014). On suppose que les mammifères ne possédant pas de Z1 (porc, chat, souris, rat) l'ont perdu de façon secondaire au cours de l'évolution.

La comparaison des séquences d'A3 de différentes espèces de primates montre que la majorité des mutations sont des substitutions (Sawyer, Emerman, & Malik, 2004). Il a été montré pour *A3G* que la plupart ces substitutions sont des mutations non synonymes. Ces mutations sont le résultat de fortes pressions de sélection positive. L'une des hypothèses serait que les A3 jouent un rôle en tant que facteurs de restriction contre les rétrovirus endogènes afin de préserver la stabilité génomique.

D'autres A3 ont évolué vers une atténuation de leur activité. C'est le cas d'*A3DE*. Chez l'Homme, *A3DE* ne possède pas d'activité. Des études ont montré qu'une mutation d'une cystéine en tyrosine en position 320 suffisait à restaurer son activité (Dang et al., 2011). Cette forme 320Y n'est retrouvée que chez 0,1% de la population humaine d'après 1000 Genomes. Il s'agit en revanche du phénotype dominant chez certains primates tels que le macaque et l'Orang Outan. Nous nous sommes intéressés à *A3DE* et avons déterminé la fonction d'*A3DE* 320C. Nous développerons cet aspect plus en détail dans les résultats.

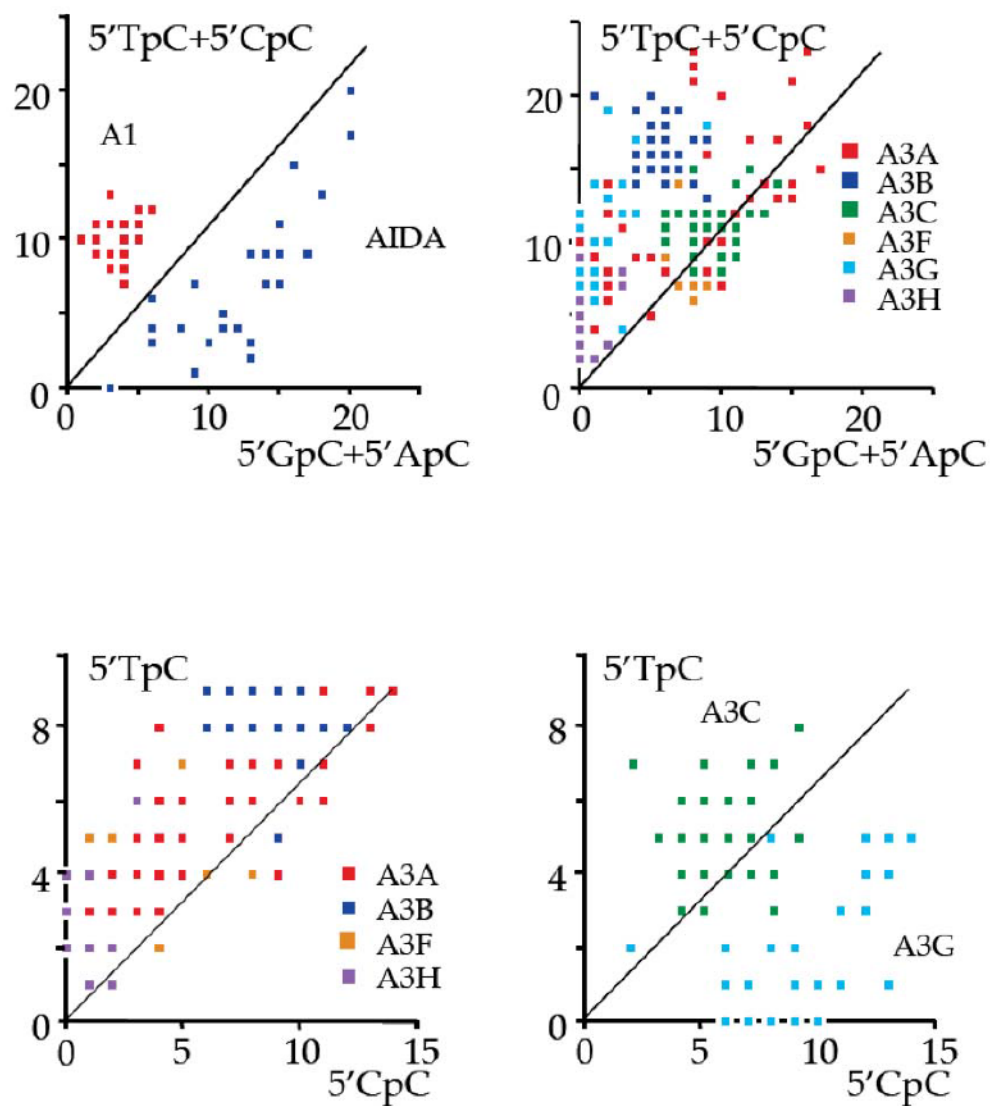


Figure 12 : Contexte de nucléotidique de désamination des différentes A3. Les contextes ont été analysés sur des génomes de VHB in vitro désaminés par les différentes A3. D'après Vartanian et al, 2010.

A3B est une autre A3 dont l'activité s'est atténuée au cours de l'évolution. En effet, l'analyse des séquences en acides aminés montre que la partie carboxy-terminale d'A3B dérive d'A3A. Bien qu'A3B soit active sur l'ADN nucléaire, celle-ci est moins efficace qu'A3A et demeure incapable d'induire des cassures double brin. Cette atténuation est apparue dans l'évolution il y a environ 38 millions d'années, suite à la divergence entre les Hominoïdes (ou grands singes) et les cercopithécidés. Nous nous sommes intéressés à cette atténuation et en avons déterminé les mécanismes moléculaires. Cette étude et ses conclusions seront développées dans la partie résultats.

viii. Signature mutationnelle des AID/APOBEC

Des études *in vitro* chez la bactérie *E. Coli* et sur le génome du Virus de l'Hépatite B ont permis de mettre en évidence que ces cytidines désaminases ont la particularité d'induire les mutations dans un contexte dinucléotidique particulier (Beale et al., 2004; Suspène, Guétard, et al., 2005; Vartanian et al., 2010). Le contexte dinucléotidique est défini par la base mutée et la base se trouvant en amont de celle-ci. Ainsi, pour une mutation touchant une cytidine précédée d'une guanosine, le contexte dinucléotidique sera noté 5'GpC, le « p » symbolisant la liaison phosphate entre la guanosine et la cytidine. Ce contexte peut différer selon la cytidine désaminase AID/APOBEC. Ceci va avoir une importance dans l'étude des mutations. En effet, les expériences *in vitro* montrent qu'A1 par exemple, mute préférentiellement dans le contexte 5'TpC. Les A3, à l'exception d'A3G, ont une préférence pour le contexte 5'TpC (Figure 12). AID en revanche induit des mutations dans le contexte 5'GpC mais aussi 5'ApC (Figure 12). Le contexte dinucléotidique va servir de signature mutationnelle : son étude permet donc de déterminer si les mutations observées sont imputables à une cytidine désaminase. L'étude des signatures mutationnelles a ainsi permis de mettre en évidence l'impact d'AID et d'APOBEC dans de nombreux processus impliquant la déamination de cytidines tels que la commutation de classe des IgG, les hypermutations somatiques, la mutation de génomes viraux et l'apparition de cancers.

II. APOBEC et virus

A. Historique

Avant de connaître le rôle antiviral des A3, de nombreuses études ont montré la présence de mutations de type GC->AT, extensives et monotones, qualifiées d'hypermutations, dans les génomes viraux. Ceci est le cas pour les rétrovirus tel que le Virus de l'Immunodéficience Humaine de type 1 (HIV-1) et de type 2 (HIV-2) (Gao et al., 1992; Goodenow et al., 1989; Vartanian, Meyerhans, Asjo, & Wain-Hobson, 1991), le Virus de l'Immunodéficience Simienne (SIV) (Johnson, Hamm, Goldstein, Kitov, & Hirsch, 1991), le Virus de l'Encéphalite Caprine (CAEV) (Wain-Hobson, Sonigo, Guyader, Gazit, & Henry, 1995), l'Anémie infectieuse équine (EIAV) (Zielonka et al., 2009), le Virus T-Lymphotrope Humain (HTLV) (Vartanian et al., 1991) ou le virus Foamy (Lochelt et al., 2005). Des mutations de types GC->AT sont aussi retrouvées sur les génomes de virus à ADN tels que le PapillomaVirus Humain (HPV) (Vartanian, Guétard, Henry, & Wain-Hobson, 2008), l'Herpes Simplex Virus (HSV) (Suspène, Aynaud, Koch, et al., 2011), le Virus Epstein Barr (EBV) (Suspène, Aynaud, Koch, et al., 2011), le Toque Teno Virus (TTV) (Tsuge et al., 2010), ou encore les hepadnavirus tels que le Virus de l'Hépatite B (VHB) (Günther et al., 1997). Ces virus possèdent une étape simple brin au moment de leur réplication (virus à ADN et Rétrovirus) et éventuellement au moment de la transcription (virus à ADN), ce qui rend possible l'*editing* par les A3. L'implication d'une A3 dans un mécanisme de défense antiviral a été montrée pour la première fois sur le VIH avec A3G (sous le nom de CEM15) (Sheehy et al., 2002). Il a ensuite été montré qu'A3G ainsi que d'autres A3 sont inductibles par la réponse immunitaire innée. Les réponses immunitaires innée et adaptative sont initiées par l'organisme en réponse à une infection afin de détruire le pathogène.

L'un des mécanismes de la réponse immunitaire innée lors d'une infection virale est la production d'interféron de type I qui activent de nombreuses cascades de signalisation.

Au cours de cette partie, nous détaillerons comment le système immunitaire induit la production d'A3 en réponse à l'infection virale et la façon dont les A3 restreignent leurs cibles.

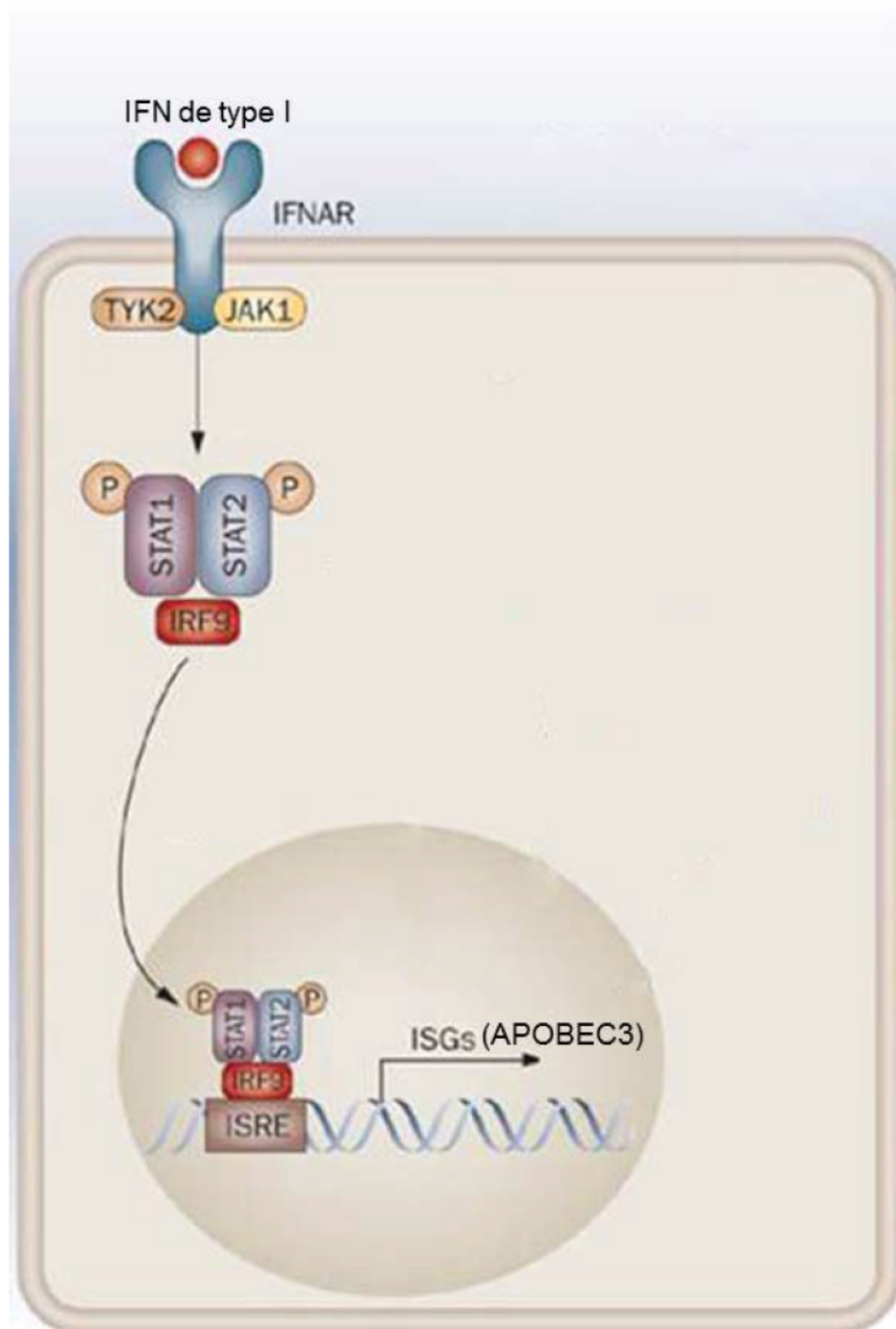


Figure 13 : Schéma du mécanisme d'induction de l'expression des A3 par l'IFN de type I. Adapté de Hall et al, 2010.

B. Réponse immunitaire et APOBEC

Lors d'une infection virale, les principales cytokines produites sont les Interférons (IFN). Ce sont des glycoprotéines secrétées par les cellules infectées. Il existe 3 types d'interférons : les interférons de type I, II et III. La principale cytokine produite lors d'une infection virale est l'IFN-I. A ce jour, 4 IFN I ont été décrits chez l'Homme : IFN- α , IFN- β , IFN- κ , et IFN- Ω . Néanmoins, seule les IFN alpha et beta semblent entrainer l'expression d'A3.

L'IFN-I a un fort effet de protection contre les infections virales et a un rôle immunomodulateur et antitumoral.

Lors de l'infection virale, les récepteurs de type Toll Like (TLR) vont reconnaître certains déterminants viraux (glycolipides, ARN, ADN simple brin, etc) et déclencher une cascade de réaction aboutissant à la production et à la sécrétion d'IFN alpha ou beta. L'IFN ainsi sécrété peut avoir un effet autocrine ou paracrine en se fixant sur les récepteurs IFNAR 1 et 2. S'ensuit une activation de la voie JAK-STAT aboutissant à la transcription des Gènes Stimulés par l'Interféron (ISG) (Figure 13). On retrouve parmi les ISG des facteurs de restriction tels que Mx, TRIM5 α ou encore PKs. Les récepteurs aux IFN α sont assez ubiquitaires, ce qui permet une réponse antivirale dans de nombreux types cellulaires.

Les premières études du rôle de l'IFN-I sur l'expression des A3 ont été faites sur les PBMC. Dans ces cellules, A3A est très fortement surexprimée et, dans une moindre mesure A3C, A3F et A3G (Refsland et al., 2010). Il a été montré qu'A3A est particulièrement sensible à l'IFN dans les lymphocytes T CD4 (Koning et al., 2009). A3A et A3G sont aussi surexprimées par l'interféron dans les monocytes et les macrophages (K. Chen et al., 2006; Koning et al., 2009; Peng et al., 2007). Les IFN beta et gamma sont aussi des inducteurs d'A3G dans les macrophages. Dans les cellules dendritiques plasmacytoides (pDC), l'IFN alpha va entrainer la surexpression d'A3A, A3C, A3F et A3G. Néanmoins, l'effet de l'IFN ne se limite pas aux cellules de l'immunité.

En effet, l'expression des A3 sous l'action de l'IFN de type I a aussi été étudiée sur les cellules hépatiques. Il a été montré que dans ces cellules, A3B, A3F et A3G sont surexprimées (Bonvin et al., 2006). Cette différence dans les A3 exprimées avec les cellules de l'immunité peut s'expliquer en partie par la nature des virus infectant ces deux types cellulaires.

Plus récemment, il a été montré qu'A3A, A3F et A3G sont aussi inductibles par l'IFN beta dans les cellules W12, lignées de kératinocytes cervicaux issus de lésions de bas grade (Z. Wang et al., 2014).

A3F et A3G possèdent un élément de réponse à la stimulation interféron (ISRE), identique, en -47 de leur codon d'initiation. A3A quant à elle possède aussi un ISRE hypothétique en -1787 de son codon d'initiation. Ceci permet d'expliquer la façon dont ces A3 sont surexprimées dans les cellules stimulées par l'interféron. Néanmoins, peu de choses sont connues sur la régulation d'A3B et d'A3C. A3DE et A3H sont les deux seules A3 dont l'expression ne dépend pas de l'interféron, bien qu'elles soient surexprimées en conditions d'inflammation dans les cirrhoses infectées par le virus de l'Hépatite B par exemple (Vartanian et al., 2010). Ceci laisse supposer un rôle particulier pour ces deux A3.

L'induction de l'expression d'A3 différentes va permettre à la cellule de se défendre contre plusieurs types de virus : en effet, les A3 vont éditer des virus différents en fonction de leur localisation cellulaire ou du cycle réplcatif du virus par exemple.

C. APOBEC et Rétrovirus « complexes »

Les rétrovirus sont divisés en deux grandes catégories : les rétrovirus « simples », dont le génome est composé des gènes canoniques des rétrovirus (gag, pol et env) et les rétrovirus « complexes » qui possèdent en plus des gènes codant des protéines dites « accessoires ». Ce sont des virus enveloppés dont le génome est composé d'un dimère d'ARN simple brin positif. Au cours de son cycle de réplication, le génome va être retrotranscrit grâce à une ADN polymérase ARN-dépendante virale. C'est au court de ce processus que les A3 vont éditer le génome viral. L'ADN obtenu va être intégré au génome de la cellule et servira de matrice pour la transcription d'ARN viraux. Ceux-ci serviront à la production des protéines virales à l'issue de la traduction, ou de génome viraux lors de l'assemblage des virons. Parmi les rétrovirus complexes, on retrouve le VIH, HTLV et le virus Foamy ; tous trois sont la cible des A3.

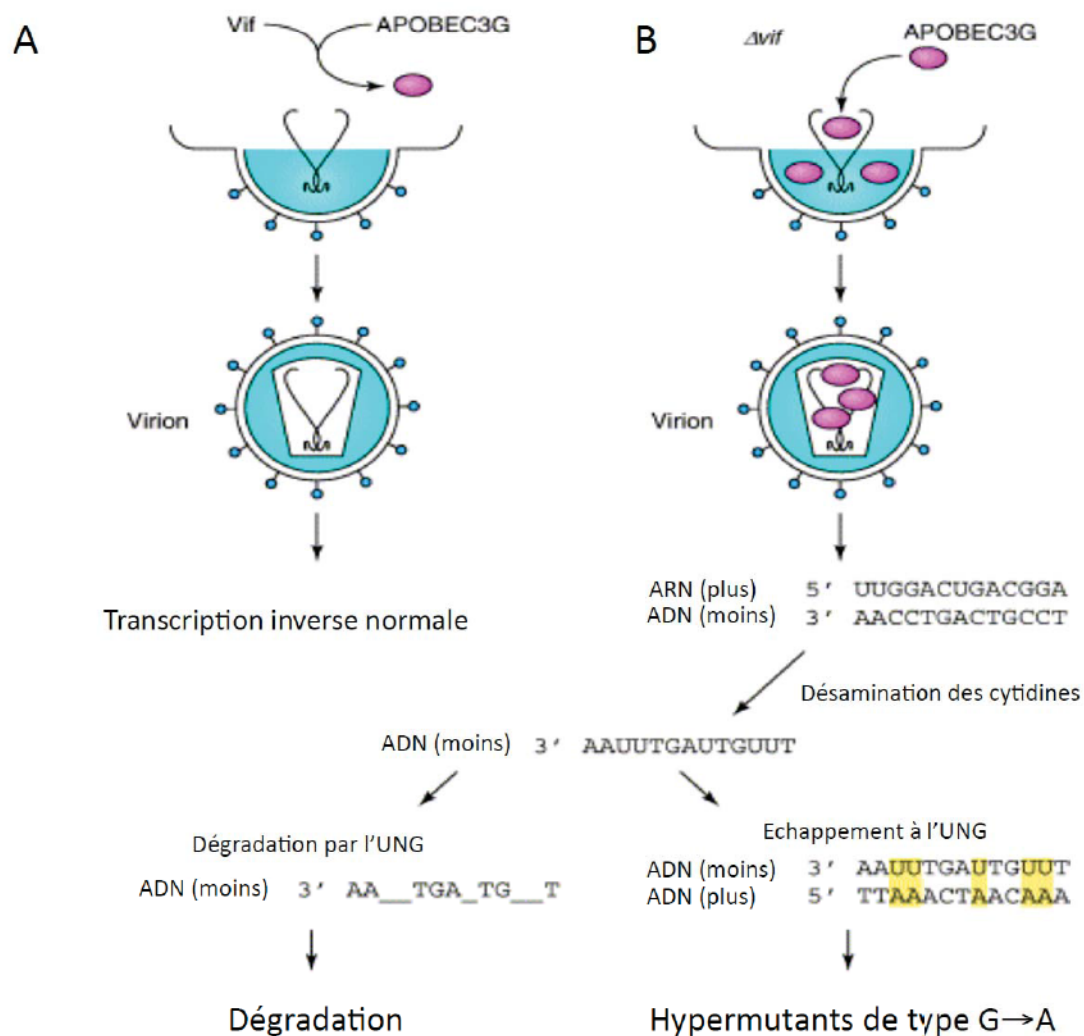


Figure 14 : Schéma de la restriction du VIH par A3G. A) La protéine Vif empêche l'encapsidation d'A3G. B) En absence de Vif, A3G est encapsidée dans les virions. Lorsque ces virions infectent de nouvelles cellules cibles, A3G entraîne la désamination des cytidines sur l'ADN viral néosynthétisé (brin -). Ces génomes mutés peuvent ensuite être dégradés par l'UNG ou alors échapper à cette dégradation donnant ainsi naissance à des génomes portant des hypermutations G→A. D'après Vartanian et al, 2003.

1. Le Virus de l'Immunodéficience Humaine

Le VIH est un *Retroviridae* appartenant à la sous-famille des *Orthoretroviridae* et au genre des *Lentivirus*. Son génome est un dimère d'ARN simple brin positif long de 9,75 kb. Le HIV est transmis par voie sanguine, sexuelle ou lors de l'allaitement. Il est responsable du Syndrome d'Immunodéficience Acquise (SIDA). Ce virus infecte principalement les cellules immunitaires telles que les lymphocytes T CD4 et les macrophages. Le cycle réplcatif se déroule en deux temps : une phase précoce et une phase tardive. La phase précoce correspond à la fusion de l'enveloppe du virus avec la membrane plasmique de la cellule cible et à la libération de la capsid. Cette capsid va ensuite être acheminée au noyau : durant le transport, l'ARN viral est retrotranscrit en ADN viral qui sera intégré au génome de la cellule. C'est lors de cette rétrotranscription que les A3 vont induire des mutations sur le génome viral. Etant donné que la rétrotranscription se déroule dans la capsid, il faut que les A3 aient été préalablement encapsidées lors d'un précédent cycle viral pour éditer l'ADN viral. Les A3 vont induire des mutations de type C->U sur l'ADN néosynthétisé : c'est mutations peuvent conduire au blocage de la réplication si elles sont trop nombreuses. Lors de l'import dans le noyau, celles-ci seront corrigées par la machinerie cellulaire aboutissant à des mutations C->T ou G->A (Figure 14). A3F et A3G restreignent aussi le virus par inhibition de la polymérase virale ou de l'intégrase.

L'intégration de l'ADN viral sous forme de provirus permet la production d'ARNm viraux, ce qui marque l'entrée dans la phase tardive de la réplication. Les ARNm viraux sont exportés dans le cytoplasme et traduits en protéines virales. Parmi elles, la protéine Vif (pour Viral Infectivity Factor), qui a pour rôle d'empêcher l'encapsidation d'A3G et d'A3F. Celle-ci interagit avec la partie amino-terminale d'A3F ou d'A3G, ce qui a pour effet d'exclure directement la protéine de la capsid. Néanmoins, il ne s'agit pas du mécanisme principal : en effet, Vif peut aussi induire la dégradation de ces A3 par le protéasome (Marin, Rose, Kozak, & Kabat, 2003; Sheehy, Gaddis, & Malim, 2003; Stopak, de Noronha, Yonemoto, & Greene, 2003). Pour cela, Vif forme un complexe avec les protéines elongin B et C qui recrutent RBX1 et Culin 5. La protéine Culin 5 lie la E2 ubiquitin conjugating qui va ubiquitiner A3F ou A3G et entrainer leur dégradation par le protéasome (Figure 15). La protéine Vif est donc une protéine fondamentale pour la survie du virus. Lorsqu'il y a encapsidation d'A3F ou

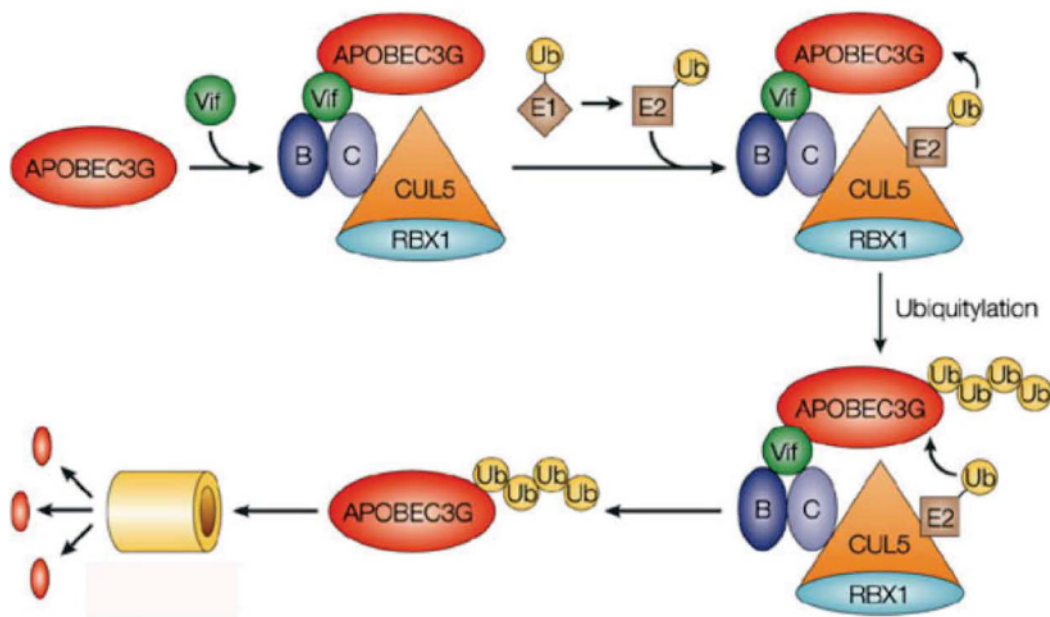


Figure 15 : Mécanisme de dégradation d'A3G par la protéine Vif. Vif se lie à APOBEC3G et recrute l'elongin B et C pour former le complexe E3 ubiquitine ligase. Les elongins vont recruter les protéines cullin 5 et RBX1, permettant à cullin 5 de recruter l'E2 ubiquitin-conjugating qui ubiquitine A3G. S'ensuit la dégradation d'A3G par le protéasome. D'après Harris et al, 2004.

d'A3G, celle-ci se fait par interaction de l'A3 avec le domaine de la nucléocapside de la protéine gag ou avec l'ARN viral : c'est le cas notamment pour les virus VIH déficients pour la protéine Vif.

Des études plus récentes tendent à montrer que certains des haplotypes d'A3H (haplotypes II, V et VII) sont capables d'induire des hypermutations sur le génome viral. De façon intéressante, on observe des mutations sur Vif permettant la dégradation d'A3H dans les régions du monde où les haplotypes d'A3H stables sont les plus fréquents. Il semblerait aussi que l'haplotype d'A3H, dans la population Japonaise, soit corrélé à la susceptibilité à l'infection HIV et à la progression vers le SIDA (Sakurai et al., 2015).

A3DE est aussi considérée comme un acteur de restriction du VIFet est aussi dégradé par Vif. Néanmoins, le faible taux de mutation obtenu avec A3DE semble indiquer que l'inhibition par A3DE passe par un mécanisme indépendant de l'activité cytidine désaminase (Dang et al., 2011; Dang, Wang, Esselman, & Zheng, 2006).

A3A agit aussi en tant que facteur de restriction dans les cellules de la lignée myéloïde (Berger et al., 2011). Elle y constitue un facteur de restriction important car elle n'est pas dégradée par Vif. Néanmoins, elle est partiellement inhibée par la protéine accessoire Vpx du HIV-2 et du SIV.

2. Le Virus T-Lymphotrope Humain

Le virus HTLV est un *Retroviridae* de la sous famille *Orthoretrovirinae* et de la famille *Deltaretrovirus*. Son génome est un ARN positif de 8,5kb maintenu sous forme de dimère dans la capside virale. Le tropisme du virus varie selon l'espèce : lymphocytes T CD4 pour le HTLV-1, CD8 pour le HTLV-2 et mixte pour le HTLV-3. La transmission peut être sexuelle ou avoir lieu lors de l'allaitement. Le HTLV-1 peut provoquer des lymphomes T de l'adulte (ATL) bien que l'infection soit dans la majorité des cas asymptomatique. Il peut aussi être responsable de neuromyélopathies chroniques appelées paraparésie spastique tropicale (PST). Il s'agit du premier rétrovirus oncogène décrit.

La proximité phylogénique entre HIV et HTLV et le tropisme similaire de ces deux virus pour les CD4 ont amené différentes équipes à étudier les rôles des A3 dans la défense contre l'infection par HTLV-1. De façon contradictoire, certaines équipes montrent qu'A3G est

incorporée dans la capside du HTLV-1 et inhibe son infectivité sans pour autant induire des mutations G->A (Mahieux et al., 2005; Sasada et al., 2005), tandis que d'autres équipes montrent que le HTLV-1 est insensible à la restriction par les A3 (Ohsugi & Koito, 2007). Il semblerait qu'A3G, mais aussi A3B, A3C et A3F soient en mesure d'éditer le génome du HTLV-1 mais qu'il s'agisse d'un événement assez rare *in vitro* pour qu'il soit difficile de le mettre en évidence *in vivo*. Plus récemment, il a été montré qu'A3A, A3B et l'haplotype II d'A3H sont encapsidés et inhibent le HTLV-1 et que cette inhibition est indépendante de l'activité cytidine désaminase pour A3H-II (Ooms, Krikoni, Kress, Simon, & Munk, 2012). Enfin, l'analyse de séquences de provirus HTLV-1 issus de patients infectés montrent que l'on retrouve des mutations G->A dans un contexte dinucléotidique correspondant à A3G (Fan et al., 2010).

Contrairement au VIH, HTLV ne possède pas de protéine accessoire entraînant la dégradation des A3. De façon intéressante, la comparaison des séquences d'acide aminé des nucléocapsides (NC) de VIH-1 et HTLV-1 montre que la NC du HTLV-1 possède une séquence plus longue d'une trentaine d'acides aminés et riche en leucine en position C terminale. Des délétions et des mutations dans cette région améliorent l'encapsidation d'A3G et entraînent une augmentation du nombre de mutations sur le génome du HTLV-1 : la NC joue donc un rôle anti A3 chez ce virus. Le HTLV-1 a donc développé un mécanisme différent du HIV au cours de l'évolution pour échapper à la restriction par les A3 (Derse, Hill, Princler, Lloyd, & Heidecker, 2007).

3. Le virus Foamy

Le virus Foamy est un *Retroviridae* de la sous famille des *Spumaretrovirinae* et du genre *Spumavirus*. Son génome est long de 12kb environs. Ces virus infectent aussi bien les félins (Feline Foamy Virus, FFV), les bovins (Bovine Foamy Virus, BFV) et les primates (Simian Foamy Virus). La transmission à l'Homme se fait de façon zoonotique, du singe à l'Homme, souvent par la salive (morsure). Le virus foamy n'est associé à aucune pathologie et l'infection est asymptomatique et persistante à vie. Son cycle réplcatif diffère de celui du HIV du fait que la retrotranscription peut avoir lieu lors de la phase tardive, aboutissant à la

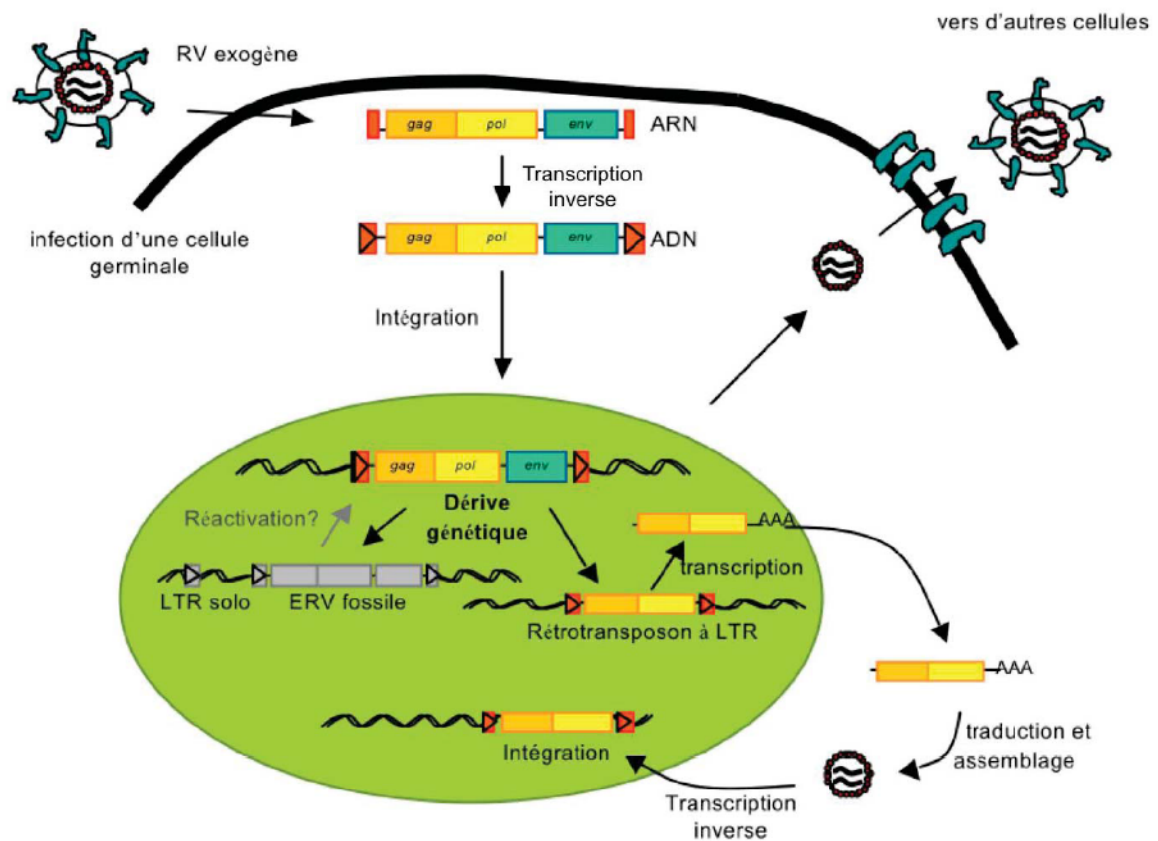


Figure 16 : Les rétrovirus endogènes. L'endogénisation des rétrovirus exogènes est le résultat de l'infection de la lignée germinale. L'accumulation de mutations et de délétions sur le provirus aboutit à son inactivation (ERV fossiles) ou à une évolution en rétrotransposons LTR (comme IAP ou MusD) qui se répliquent par un cycle intracellulaire.

présence du génome viral dans la capside sous forme d'ADN double brin au lieu d'un dimère d'ARN négatif simple brin.

Les virus foamy sont aussi la cible de la restriction par les A3 dans les différentes espèces qu'ils infectent. Il a été montré que l'A3G humaine, et dans une moindre mesure A3B et A3F, inhibent la réplication du virus (Delebecque et al., 2006). A l'instar du VIH, le virus foamy possède une protéine accessoire, Bet, qui a pour rôle de contrecarrer l'action des A3. Contrairement à Vif, Bet n'entraîne pas la dégradation d'A3G mais semble plutôt empêcher son encapsidation en inhibant sa dimérisation pour entraîner la formation d'oligomères d'A3G et de Bet, diminuant ainsi la solubilité cytosolique d'A3G (Russell et al., 2005). La protéine Bet interagit aussi avec A3C, au niveau de son site d'homodimérisation, ce qui a pour effet d'empêcher la dimérisation et ainsi d'inhiber son activité cytidine désaminase (Perkovic et al., 2009). Par ailleurs, la protéine Bet du FFV interagit aussi avec l'A3 féline et inhibe son activité de restriction (Lochelt et al., 2005).

4. Les Rétrovirus endogènes

Au cours de l'évolution, certains rétrovirus se sont maintenus sous forme intégrée dans le génome de l'hôte, au niveau de la lignée germinale, permettant ainsi une transmission mendélienne du génome viral. Ces rétrovirus sont appelés rétrovirus endogènes (ERV). Il existe deux grandes catégories d'ERV : les rétroéléments sans LTR (SINEs et LINEs) et avec LTR (Rétrotransposons à LTR). Ces derniers sont caractérisés par leur séquence contenant des gènes *gag*, *pol* et *env* fossiles et flanqués de séquences LTR. La majorité des ERV possèdent un génome rétroviral simple, composé des gènes *gag*, *pol* et *env*. Néanmoins, il existe des exceptions telles que l'ERV Humain K (HERV-K) qui possède aussi le gène *Rec*, un homologue de la protéine Rev du VIH. Les ERV seraient le résultat de l'infection de cellules de la lignée germinale par des rétrovirus : ceux-ci se seraient alors maintenus dans le génome de l'hôte à la suite d'une dégénérescence de leur séquence abolissant leur virulence. Ces événements d'endogénisation datent de plusieurs millions d'années (Figure 16). On estime par exemple l'endogénisation de HERV-K ancienne d'environ 30 millions d'années. Ces ERV peuvent avoir un impact négatif sur la stabilité génomique de l'hôte mais aussi

contribuer à son évolution comme c'est le cas pour HERV-W impliqué dans la mise en place du placenta.

Chez l'Homme, près de 8% du génome sont constitués de HERV. Contrairement aux rétrovirus « actifs », les HERV s'intègrent principalement dans des régions non actives ou très peu actives transcriptionnellement ou dans des introns, en position antisens de gènes actifs transcriptionnellement. Ceci est le résultat d'une sélection naturelle favorisant les virus intégrés les moins actifs et donc aux effets moins nocifs pour la cellule. De plus, plusieurs mécanismes de régulation épigénétiques tel que la méthylation et la phosphorylation des histones sont mis en place par la cellule afin de réguler au maximum l'activité de ces HERV.

Il a été montré que les A3 eux aussi participent au contrôle de l'activité des ERV. Les premières études se sont d'abord intéressées aux ERV à LTR de levure (Ty1) et de souris (IAP et MusD) dont la structure est proche des HERV à LTR (Esnault et al., 2005; Schumacher, Nissley, & Harris, 2005). Il a d'abord été montré qu'A3C, A3G et A3F sont capables d'inhiber le rétrotransposon Ty1 chez la levure en éditant son ADN complémentaire (cDNA). A3A, A3B et A3G, peuvent, quant à elles, inhiber le rétrotransposon IAP (Esnault, Priet, Ribet, Heidmann, & Heidmann, 2008) : A3B et A3G sont encapsidées par interaction avec la protéine gag du IAP et induisent ensuite des mutations sur le génome de celui-ci. Ce n'est pas le cas pour A3A dont le mécanisme d'inhibition reste à déterminer. A3G est aussi active sur le rétroélément murin MusD. Les rétrotransposons sans LTR peuvent aussi être la cible des A3. C'est le cas de LINE-1 qui est inhibé par toutes les A3 avec une efficacité et un mécanisme parfois différents. En effet, A3A, A3B, A3F et A3G inhibent fortement LINE-1 tandis qu'A3C inhibe modérément LINE-1 et A3DE et A3H faiblement (Kinomoto et al., 2007; Muckenfuss et al., 2006; OhAinle et al., 2008). A3G inhibe aussi les éléments SINE en séquestrant ses ARN dans des complexes ribonucléotidiques.

De manière intéressante, on observe que l'augmentation du nombre d'A3 chez les primates est corrélée à l'activité des rétroéléments. En effet, seuls les rétroéléments sans LTR (LINE et SINE) sont actifs chez les primates tandis que chez la souris, qui ne possède que 3 APOBEC, des rétroéléments avec et sans LTR sont actifs. On estime qu'il y a 50 à 60 fois plus de LINE actifs chez la souris comparés à l'Homme, et que ceux-ci sont impliqués dans 35% de plus de maladies que chez l'Homme (Schumann, 2007).

Une hypothèse propose que le maintien de la stabilité génomique ait constitué une pression de sélection aboutissant à l'expansion des A3, permettant ainsi de réguler l'activité des rétrotransposons.

La découverte de mutations G->A imputables à A3G fixées dans les séquences de rétrovirus endogènes de la famille HERV-K soutient cette hypothèse (Armitage et al., 2008).

D. APOBEC3 et Virus à ADN

Les A3 sont aussi impliquées dans la réponse immunitaire contre les virus à ADN. Ces virus possèdent un génome sous forme d'ADN double brin ou simple brin. Pour se répliquer, ceux-ci acheminent leur génome dans le noyau. C'est à ce niveau que les virus à ADN double brin vont être mutés par les A3 : en effet, au cours de la réplication et de la transcription, leur génome se retrouvera sous forme simple brin et donc accessible aux A3. Parmi les virus à ADN double brin inhibés par les A3, on retrouve le PapillomaVirus Humain (HPV), le Virus de l'Herpes Simplex (HSV) et le Virus Epstein-Barr (EBV). Le seul virus à ADN simple brin inhibé par les A3 connu à ce jour est le Virus Adéno-Associé (AAV). Nous allons détailler le mécanisme de restriction pour chacun de ces virus.

1. Les virus à ADN double brin

a. Les papillomavirus humains

Les papillomavirus humains sont des virus à ADN double brin circulaire d'une taille de 8kb. Ils appartiennent à la famille des *Papillomaviridae* et au genre *Papillomavirus*. Il existe près de 200 génotypes mais seuls 120 génotypes ont été identifiés aujourd'hui. Ces virus peuvent avoir un tropisme pour l'épithélium génital ou cutané selon le génotype. Le cycle réplcatif du HPV sera différent selon l'état de différenciation de la cellule. Le HPV infecte la couche basale de l'épithélium. Le génome viral est ensuite libéré dans le noyau et y

persiste sous forme d'épisome. Lors de la division cellulaire, le génome viral est répliqué au même titre que l'ADN génomique de la cellule hôte par la machinerie cellulaire. Au cours de cette phase, les gènes de la phase précoce (Early) sont transcrits et traduits. Lorsque la cellule infectée se différencie en kératinocyte, la transcription et la traduction des gènes tardifs (Late), qui codent les protéines structurales, sont initiées. Ceci s'accompagne d'une très importante réplication de l'ADN viral, étape essentielle à la production virale étant donné que les cellules différenciées ne se divisent plus, et ne répliquent donc plus d'ADN. Ce cycle va permettre la production de virions. Pour de nombreux génotype, ceci va s'accompagner de lésion de la peau (verrues, condylomes) ou de la muqueuse bénignes (lésions de bas grade). Néanmoins, d'autres génotypes sont associés à l'apparition de cancers cutanés ou muqueux tels que le cancer du col de l'utérus (lésion de haut grade).

Les kératinocytes sont des cellules exprimant A3A et A3B (Madsen et al., 1999; Trivedi, Gilliland, Zhao, Liu, & Thiboutot, 2006). De plus, ces deux A3 ont une localisation nucléaire. Ceci laisse donc supposer qu'elles peuvent éditer le HPV. Ceci a été montré pour la première fois en 2008 : dans cette étude, les auteurs ont pu mettre en évidence des séquences de HPV-1a et HPV16 hypermutés par A3, *in vivo*, avec une fréquence de 11% et 9% respectivement (Vartanian et al., 2008). Cette étude apporte aussi la preuve qu'A3A, A3C et A3H, toutes trois localisées dans le noyau, peuvent éditer le génome de HPV-1a et HPV-16 contrairement aux autres A3. Plus récemment, il a été montré *in vitro* qu'A3A et A3G, dont la surexpression est induite par l'interféron bêta, peuvent provoquer des mutations sur le génome du HPV-16 et plus particulièrement sur le gène *E2*. Cette observation est confirmée *in vivo*, où les auteurs retrouvent des mutations de signature A3 sur E2, dans des néoplasies du col de l'utérus infectées par HPV. Par ailleurs, l'étude des mutations de deux cancers causés par HPV (cancer du col de l'utérus et des voies aérodigestives supérieures) met en évidence un nombre important de mutations imputables aux A3, corrélé avec une surexpression d'A3A et A3B. Cette étude montre qu'A3A inhibe fortement l'ineffectivité des virions assemblés en présence d'A3A *in vitro* et que, dans ce système expérimental, la protéine E7 est impliquée dans un mécanisme aboutissant à la surexpression d'A3A et A3B (Z. Wang et al., 2014; Warren et al., 2014).

b. Les Herpesvirus

Les *herpesviridae* sont des virus à ADN double brin linéaire d'une taille allant de 120 à 240 kb. Ceux-ci forment une famille rassemblant trois sous familles : les Alphaherpesvirinae dont l'un des représentant est le virus de l'herpès simplex 1 (HSV-1), les Betaherpesvirinae dont l'un des membres est le cytomegalovirus (CMV) et enfin les Gammaherpesvirinae dont fait partie, entre autre, l'EBV. L'impact des A3 sur la réplication des herpesviridae n'a été étudié que dans le cas du HSV-1 et de l'EBV. Tandis que le HSV-1 a un tropisme cellulaire très large, l'EBV infecte principalement les lymphocytes B et les cellules épithéliales. A l'instar du HPV, ces deux virus acheminent leur ADN vers le noyau qui sera le lieu de réplication et de transcription de l'ADN viral. Ceci implique une exposition de l'ADN viral aux A3 nucléaires.

Il a été montré *in vitro* qu'A3A, A3C, AID et A3G peuvent éditer le génome du HSV-1 et de l'EBV mais que seule A3C inhibait efficacement l'infectivité de ces virus. De plus, des mutations dont la signature correspond aux A3 sont retrouvées *in vivo*, sur des prélèvements de lésions buccales (Suspène, Aynaud, Koch, et al., 2011).

Le tropisme de l'EBV pour les lymphocytes B pose la question du rôle d'AID dans la restriction de ce virus *in vivo*. Néanmoins, il a été montré que la protéine EBNA-2 inhibe la production d'AID et aucune étude n'a jamais montré de restriction de l'EBV par AID (Tobollik et al., 2006).

Par ailleurs, il a été montré que A1 pouvait jouer un rôle dans la restriction du HSV-1 : en effet, l'infection par le HSV-1 est inhibée dans des lignées stables pour A1. L'inhibition de l'expression de A1 dans ces lignées permet de rétablir l'ineffectivité du HSV-1, ce qui démontre que la restriction est bien l'œuvre d'A1.

2. APOBEC 3 et virus à ADN simple brin

Les virus à ADN simple brin sont généralement des petits virus dont le génome fait moins de 10kb. Parmi les virus à ADN simple brin étudiés dans le cadre de la restriction par les APOBEC, on retrouve le AAV et le Minute Virus of Mice (MVM), virus très proche du B19 humain. Ce sont tous deux des *Parvoviridae* dont le génome est compris entre 4 et 6 kb. Tandis que le MVM, appartenant au genre *Protoparvovirus*, infecte les fibroblastes ou les lymphocytes de façon autonome, le AAV, appartenant au genre *Dependoparvovirus*, dépend de l'infection préalable de la cellule hôte par un adénovirus ou un herpesvirus. La réplication

de ces deux virus est nucléaire et commence par la synthèse du brin complémentaire du génome viral. L'ADN double brin ainsi synthétisé sert de matrice à la transcription mais aussi à la réplication en cercle roulant, permettant la synthèse de génomes viraux simple brin.

De par la nature de leur génome, ces virus ont été étudiés dans le cadre du rôle antiviral des A3. Dans une première étude, il a été montré que A3A restreint de façon efficace l'infection par le AAV. Par ailleurs, cette étude montre par microscopie une colocalisation d'A3A au niveau des centres réplicatifs du AAV dans les cellules infectés, et que cette inhibition de l'infectivité du AAV est dépendante de l'activité d'A3A par mutagenèse sur le site catalytique au niveau du résidu 106 (H. Chen et al., 2006). Néanmoins, une étude plus récente ayant comparé différents mutants catalytiques, dont le mutant F75L, montre que la restriction par A3A n'est pas dépendante de l'activité cytidine désaminase puisque ce mutant inhibe la réplication du AAV sans posséder d'activité cytidine désaminase. Par ailleurs, cette étude démontre aussi qu'A3A inhibe la réplication du parvovirus MVM, encore une fois, indépendamment de l'activité catalytique d'A3A. De plus, cette inhibition est aussi possible lorsque le gène de capsid est délété, suggérant que l'inhibition a lieu au niveau de la réplication virale dans le noyau et non lors de l'encapsidation. De plus, cette inhibition nécessite la région linker et pseudoactive d'A3A (Narvaiza et al., 2009).

3. Le virus de l'Hépatite B

Le virus de l'Hépatite B est à l'origine d'hépatites virales. A ce jour, 350 millions de personnes dans le monde sont porteurs du VHB tandis que 2 milliards d'individus ont déjà fait une infection à VHB. Dans 90% des cas, l'infection aboutit à une guérison spontanée. Néanmoins, l'infection peut devenir chronique et évoluer vers des cirrhoses ou des carcinomes hépatocellulaires. Ainsi, 15% à 25% des porteurs chroniques meurent de l'une de ces deux complications. Le taux d'infection chronique chez les enfants et les nourrissons peut atteindre 90%, notamment dans les zones endémiques telles que l'Afrique Sub-Saharienne ou l'Asie du Sud-Est, ce qui en font des populations particulièrement à risque.

Le VHB appartient à la famille des *Hepadnaviridae* et au genre *Orthohepadnavirus*. Son génome est un ADN circulaire partiellement double brin d'environ 3,2 kb.

Ce génome comporte 4 gènes : le gène pré-S/S codant les protéines de surface, le gène pré-C/C codant les protéines structurales de la capside et l'antigène HBe, le gène P codant la polymérase virale et enfin le gène X codant une protéine dont le rôle dans le cycle viral est encore inconnu. Le VHB infecte uniquement les cellules hépatiques. Le virion se fixe sur son récepteur, le récepteur au sulfate d'héparane, ce qui entraîne la fusion de l'enveloppe virale à la membrane cellulaire. Le génome viral est ensuite acheminé vers le noyau. Cet ADN, partiellement double brin, va être réparé dans le noyau de la cellule hôte et va y demeurer sous la forme d'un ADN en anneau fermé (cccDNA). Celui-ci va servir à la production d'ARN messagers viraux mais aussi à la production d'ARN pré-génomiques qui, une fois rétrotranscrits en ADN dans le cytoplasme, vont être encapsidés dans les virions néo-formés. La rétrotranscription est assurée par l'ADN polymérase ARN-dépendante virale et va aboutir à la production d'un ADN partiellement double brin. Le brin d'ADN négatif est totalement synthétisé tandis que le brin positif ne l'est que partiellement. Ceci constitue une spécificité des *Hepadnaviridae* puisque la majorité des virus à ADN utilisent la machinerie cellulaire pour la réplication de leur génome. Les virions bourgeonnent au réticulum endoplasmique (RE) et y acquièrent leur enveloppe et les protéines associées à celle-ci. Les virions suivent ainsi la voie de sécrétion en passant par l'appareil de Golgi pour enfin être sécrétés hors de la cellule.

Le cycle réplcatif du VHB laisse la possibilité d'un *editing* par les APOBEC lors de deux étapes : au moment de la transcription inverse et lors de la transcription de l'ADNccc. Les premières séquences affichant des mutations de type G->A ou C->T ont été identifiées en 1997 (Günther et al., 1997). Il a ensuite été montré que les cytidines désaminases A3G, A3B, A3F et A3C peuvent agir en tant que facteurs de restriction contre le VHB (Bonvin et al., 2006; Suspène, Guétard, et al., 2005). En effet, ces cytidines désaminases peuvent induire des mutations sur le génome du VHB au moment de la réplication. Il a été montré plus récemment qu'A3A et A3H pouvaient aussi être impliquées dans la restriction du VHB, ainsi qu'AID (Abe et al., 2009; Henry et al., 2009; Liang et al., 2013). Lors d'expériences *in vitro*, A1 est aussi capable de muter le génome du VHB (Gonzalez et al., 2009). La stimulation d'hépatocytes par l'IFN de type I entraîne une surexpression d'A3B, d'A3F et d'A3G. Cette surexpression s'accompagne d'un taux de mutations plus important sur le génome du VHB. Ces protéines, en particulier A3G, inhibent la réplication virale en induisant des mutations sur le génome viral et sont à l'origine de 35% de mutations dans le génome du VHB (Baumert,

Rosler, Malim, & von Weizsacker, 2007; Bonvin et al., 2006; Kock & Blum, 2008; Noguchi et al., 2007; Rosler et al., 2005).

Récemment, une étude s'est intéressée au mécanisme de restriction sur l'ADNccc. Il a été montré que le traitement des hépatocytes avec de l'interféron alpha ou de la lymphotoxine beta entraîne une dégradation du cccDNA consécutive à l'apparition de sites apuriques et apyrimidiques dus à des phénomènes de désamination. De plus, ce phénomène est corrélé à une surexpression d'A3A et d'A3B, ce qui amène les auteurs à montrer que cette dégradation est dépendante de l'activité d'A3A et d'A3B. Par ailleurs, ils montrent aussi qu'A3A peut interagir avec la protéine de capsid du VHB qui elle-même interagit avec le cccDNA. Ceci permet donc un accès au cccDNA pour A3A. De façon intéressante, la surexpression d'A3A et d'A3B ne semble pas d'avoir d'effet délétère sur pour la cellule dans ce modèle (Lucifora et al., 2014).

A l'instar du VIH, la restriction peut aussi être indépendante de l'activité cytidine désaminase sur le VHB. En effet, ceci a été montré pour A3B et A3G. A3B peut interagir avec la ribonucléoprotéine K, impliquée dans l'expression de certains gènes du VHB, dont le gène S, ce qui a pour effet d'inhiber la réplication du VHB (W. Zhang et al., 2008). A3G quant à elle inhibe l'étape de reverse transcription (Nguyen, Gummuluru, & Hu, 2007).

Dans l'optique d'étudier le lien éventuel entre physiopathologie (cirrhose et carcinomes hépatocellulaires) et *editing* par les cytidines désaminases, il a été montré que A3G, A3B et A3C induisent des mutations sur le gène Hbx, codant la protéine X. Ces mutations semblent être impliquées dans l'émergence de protéines X inhibant l'apoptose et conduisant à la genèse de cancers hépatocellulaires. Ceci s'accompagne d'une surexpression d'A3B (R. Xu et al., 2007).

A3G aussi pourrait aussi avoir un effet délétère sur l'hôte puisqu'elle contribue quant à elle à l'émergence de nouveaux antigènes HBe, ce qui permet au virus d'échapper au système immunitaire (Vartanian et al., 2010).

L'étude de l'expression des A3 dans des tissus cirrhotiques infectés par le VHB montre une surexpression d'A3C, A3DE, A3G et A3H (Vartanian et al., 2010).

A3DE n'étant pas active, ceci pose la question du rôle de celle-ci dans les tissus cirrhotiques infectés par le VHB. C'est ce que nous avons étudié dans la partie résultats.

Les cytidines désaminases constituent donc des puissants mutateurs des génomes viraux et sont impliquées dans la restriction de nombreux virus différents. Néanmoins, ceci implique que leur activité doit être régulée afin de ne pas avoir d'effet délétère sur le génome cellulaire et aboutir à des mutations entraînant l'apparition de tumeurs. Ceci pose aussi la question du rôle des cytidines désaminases en absence d'infection. Nous aborderons ces deux points lors de la dernière partie de cette introduction par l'étude du rôle des cytidines désaminases dans le catabolisme et dans l'apparition des cancers.

III. Rôle des APOBEC dans le catabolisme et le cancer

L'inflammation permet la surexpression des A3 et la restriction virale. Néanmoins, les A3 ont pour substrat l'ADN simple brin, sans spécificité pour les génomes viraux. Ceci implique que l'ADN cellulaire, qu'il soit cytoplasmique (ADN mitochondrial) ou nucléaire (ADN génomique) peut être la cible des A3.

Ceci présente un intérêt particulier dans les contextes d'inflammation chronique, caractéristique de certains virus oncogènes (VHB, HPV), de certaines infections bactériennes (*Helicobacter pylori*) ou de certaines maladies telles que l'athérosclérose ou le sepsis.

Quel est l'impact des A3 sur l'ADN génomique et cytoplasmique dans ces conditions ? Cet aspect a été abordé par Stenglein et ses collaborateurs : ceux-ci ont montré qu'un ADN transfecté dans une lignée exprimant fortement A3A perd en stabilité. En effet, en transfectant dans cette lignée un plasmide rapporteur exprimant la Green Fluorescent Protein (GFP), ce même groupe a montré qu'A3A entraîne des mutations sur ce plasmide et que cela aboutit à la perte de la GFP (Stenglein, Burns, Li, Lengyel, & Harris, 2010). Par ailleurs, Landry et ses collaborateurs se sont intéressés à l'impact d'A3A sur l'intégrité de l'ADN génomique. Ils ont pour cela étudié la phosphorylation de l'histone γ H2AX. En effet, celui-ci est phosphorylé au niveau de son résidu 139 lorsqu'il y a formation de DSB. Cette phosphorylation, et par conséquent la présence de cassures, peut être spécifiquement détectée à l'aide d'un anticorps par microscopie à fluorescence ou cytométrie en flux. Les auteurs ont ainsi montré qu'A3A peut induire des cassures sur l'ADN génomique et que ceci pouvait entraîner un blocage du cycle cellulaire au checkpoint G1/S (Landry et al., 2011). Ces premières découvertes ont servi de base à l'étude du rôle des A3, en particulier d'A3A dans les dommages sur l'ADN cellulaire et éventuellement l'apparition des cancers.

Par ailleurs, il est maintenant connu que l'ADN cytosolique peut induire une réponse immunitaire par la réaction des senseurs à l'ADN du soi. Il est par conséquent nécessaire, afin de ne pas entrer dans une situation d'inflammation chronique, de dégrader cet ADN cytoplasmique. Les A3, notamment A3A et A3G, étant induites par l'interféron, on peut se demander quelle peut être la contribution des A3 dans le catabolisme de l'ADN.

Nous allons, le long de cette partie, détailler dans un premier temps de quelle façon les A3 peuvent être impliquées dans le catabolisme de l'ADN cellulaire en introduisant le concept d'hypermutation puis en décrivant les senseurs et les voies pouvant être impliquées dans ce mécanisme. Enfin, nous étudierons le phénomène d'hypomutation et la façon dont les A3 peuvent être impliquée dans la genèse des cancers.

A. A3 et catabolisme de l'ADN

1. Les Hypermutations

Le phénomène d'hypermutations a d'abord été associé à AID et aux hypermutations somatiques retrouvées dans les gènes V réarrangés des immunoglobulines, notamment dans les régions hypervariables. Le taux de mutation lors des phénomènes d'hypermutations somatiques peut être 10^5 à 10^6 fois plus important. Ceci a ensuite été observé sur les génomes viraux, notamment dans le cas du VIH où 91% à 100% des substitutions étaient de type G->A. Par conséquent, ces mutations ont été caractérisées d'hypermutations extensives et monotones. Par ailleurs, en moyenne 79% de ces mutations ont lieu dans le contexte 5'GpA (équivalent à 5'TpC sur le brin complémentaire) ce qui correspond au contexte d'*editing* par les A3.

Le phénomène d'hypermutations est caractéristique de l'activité des AID/APOBEC. Ce type de mutation a ensuite été retrouvé dans le génome d'autres virus, mais aussi sur des cibles cellulaires : ceci a été décrit pour la première fois *in vivo* en 2011 sur l'ADN mitochondrial. En effet, l'analyse des mutations sur le gène *MT-COI* issus de cirrhoses de patients alcooliques, infectés par le VHB, infectés par le Virus de l'Hépatite C (VHC) ou co-infectés par le VHB et le VHC montre que dans 47% des échantillons, 30 à 80% des cibles sont éditées. Ceci est confirmé dans les PBMCs et les PBMCs transformés par l'EBV. De façon intéressante, les hypermutations sur l'ADN mitochondrial sont retrouvées dans tous les compartiments des PBMC (Lymphocytes T CD4, CD8, Eosinophiles, etc) mais sont plus abondantes dans les neutrophiles. Dans tous les cas, la signature dinucléotidique correspond aux A3 (TpC et CpC) (Suspène, Aynaud, Guétard, et al., 2011).

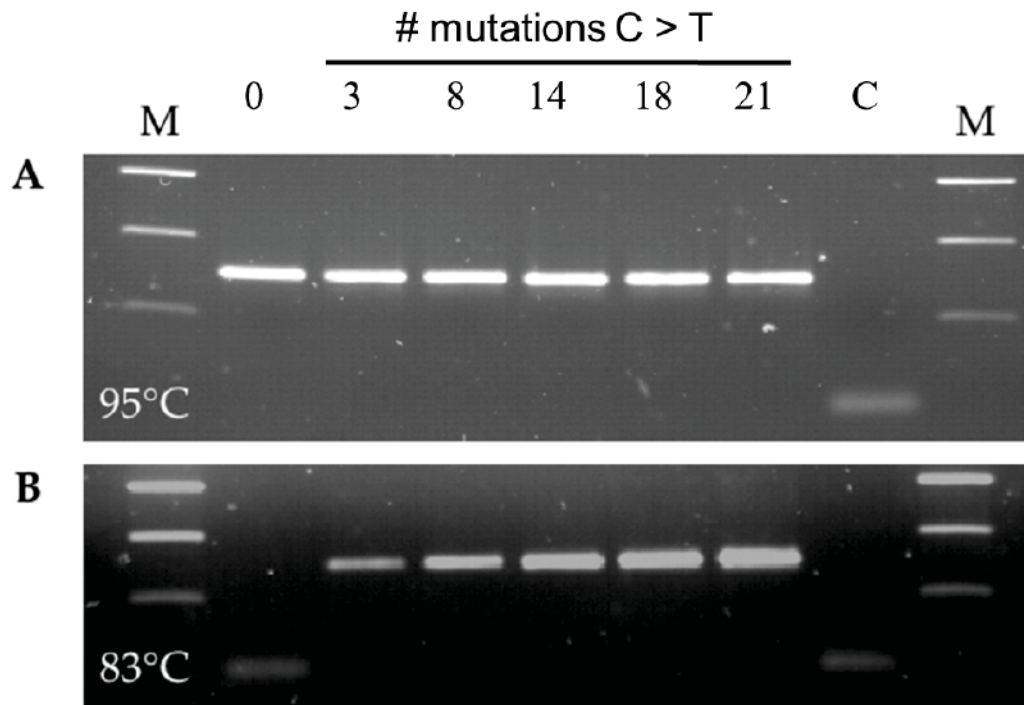


Figure 17 : Principe de la 3D-PCR. A. Amplification d'une matrice possédant 0, 3, 8, 14, 18, ou 21 mutations C->T à une température de dénaturation 95°C. Les matrices sont amplifiées avec la même efficacité quel que soit leur nombre de mutations C->T. B. Amplification d'une matrice possédant 0, 3, 8, 14, 18 ou 21 mutations C->T à une température de dénaturation de 83°C. On observe ici que, plus le nombre de mutations C->T est important, plus l'amplification est efficace. Ce phénomène traduit de la propriété de la 3D-PCR à amplifier spécifiquement les séquences hypermutés. C : contrôle négatif. M : marqueur de poids moléculaire. D'après Suspène et al. 2005.

Pour mettre en évidence les phénomènes d'hypermutations rares, la technique de 3DPCR (pour Differential DNA Denaturation-PCR) a été mise au point. Cette technique se base sur la différence entre le nombre de liaisons hydrogènes impliquées dans l'appariement des bases AT (2 liaisons hydrogènes) et GC (3 liaisons hydrogènes). Plus le nombre de liaisons est important, plus l'énergie nécessaire à la dénaturation, étape indispensable lors d'une PCR, va être importante. Ceci implique que la température nécessaire à la dénaturation d'une séquence abondante en GC va être plus importante que celle d'une séquence abondante en AT. Les cytidines désaminases induisant des mutations de type C->T ou G->A, il va y avoir une diminution du nombre de liaisons hydrogènes impliquées dans la liaison des deux brins d'ADN d'une séquence et donc une température de dénaturation nécessaire moins importante pour l'amplification de la séquence mutée comparée à la séquence sauvage. En exploitant cette caractéristique, il est possible d'amplifier spécifiquement les séquences hypermutées en abaissant la température de dénaturation à une température dite restrictive (Figure 17). Lors d'une 3DPCR, un gradient de température de dénaturation est généralement appliqué afin de visualiser la température restrictive pour chacun des échantillons. La PCR va permettre l'enrichissement en séquences hypermutées à la température restrictive et permettre leur séquençage: en effet, les hypermutations peuvent être des événements rares et ne pas être détectées par séquençage classique (Suspène, Henry, Guillot, Wain-Hobson, & Vartanian, 2005).

L'emploi de cette technique ne permet néanmoins pas de discriminer l'ADN mitochondrial encore dans la mitochondrie de l'ADN mitochondrial libre dans le cytosol. Il a été montré par microscopie confocale qu'aucune A3 n'est localisée dans la mitochondrie et donc que l'*editing* de l'ADN mitochondrial a lieu dans le cytoplasme. Cette observation implique que l'ADN mitochondrial se retrouve dans le cytoplasme. Cet ADN présente la particularité de ne pas être méthylé et est très proche de l'ADN bactérien du fait de la probable origine endosymbiotique de la mitochondrie. Cette proximité rend possible la détection de l'ADN mitochondrial cytoplasmique par la cellule et expose la cellule à un stress de source endogène. De nombreux senseurs intracellulaires ont été identifiés comme senseurs de l'ADN mitochondrial.

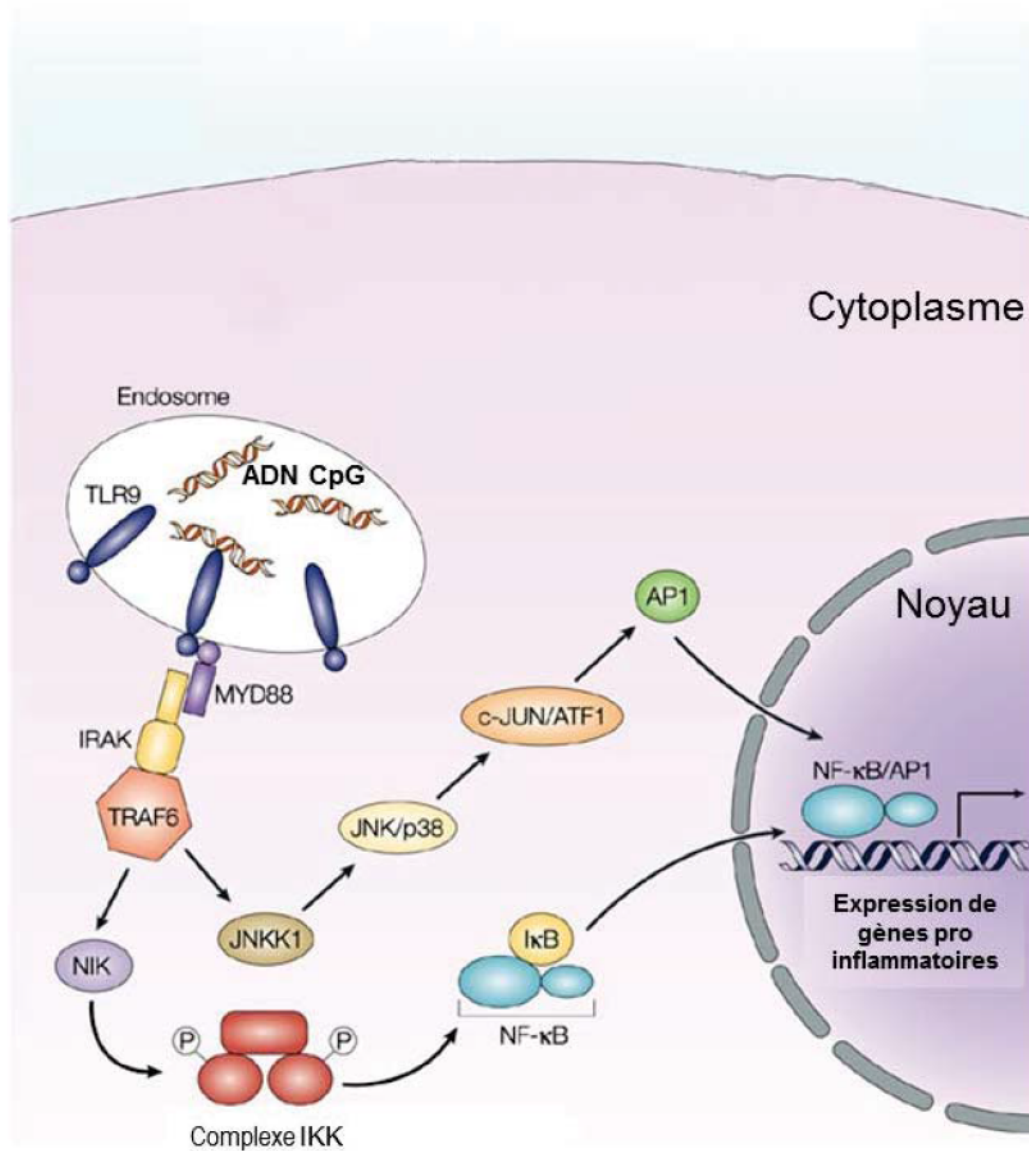


Figure 18 : Mécanisme d'induction de l'expression de gènes pro inflammatoire par l'ADN CpG par l'intermédiaire du TLR9. Adapté de Klinman et al, 2004.

2. Les censeurs de l'ADN cytoplasmique

a. La voie du Toll-Like Receptor 9

Il a été montré qu'une fuite de l'ADN mitochondrial dans le cytosol peut être détectée par le Toll-Like Receptor 9 (TLR-9) aboutissant une inflammation dans les cardiomyocytes et pouvant aboutir à des myocardies. Par ailleurs, il a aussi été montré que l'ADN mitochondrial circulant peut être à l'origine d'inflammation systémique par l'intermédiaire du TLR-9.

Le TLR-9 appartient à la famille des Toll Like Receptors. Ceux-ci sont impliqués dans le système Pathogen-associated molecular patterns/Pattern Recognition Receptors (PAMPs /PRR). Ce système fait intervenir des récepteurs (PRR) détectant des PAMPs, c'est ça dire des molécules (lipopolysaccharides, peptidoglycans, ADN possédant des CpG non méthylés, etc) essentiels à la survie des pathogènes. Ces récepteurs initient ensuite une cascade de réactions permettant la mise en place du système immunitaire inné contre ces pathogènes. L'Homme possède 13 récepteurs TLR différents (TLR-1 à TLR-13) dont cinq sont impliqués dans la reconnaissance d'acides nucléiques (TLR-3, 7, 8, 9 et 13). Ces TLR sont localisés sur les membranes des endosomes : le domaine de reconnaissance est situé dans la lumière de l'endosome tandis que le domaine de transduction du signal est cytoplasmique. Ces deux domaines sont liés par un domaine transmembranaire. Parmi les 5 TLR reconnaissant les acides nucléiques, le TLR-9 est le seul à reconnaître l'ADN. Celui-ci reconnaît spécifiquement les îlots CpG non méthylés, qui sont communs chez les bactéries mais rares chez les mammifères. La reconnaissance du ligand par TLR-9 va entraîner la dimérisation du récepteur. S'en suit une modification conformationnelle de son domaine cytoplasmique et le recrutement de la protéine MyD88. Ceci a pour effet l'activation des protéines IRAK-1 et 4.

C'est IRAK-4 qui est impliquée dans la réponse pro-inflammatoire : celle-ci va recruter TRAF6 et ainsi activer la kinase TAK-1. TAK-1 va phosphoryler le complexe IKK aboutissant à l'activation du facteur de transcription NFκB mais aussi des MAPSs kinases qui activent AP-1. NFκB et AP-1 vont permettre la transcription des cytokines pro-inflammatoires IL6, IL-12 et TNF. IRAK-1 quant à elle va permettre l'activation de IRF-7 et va agir en tant

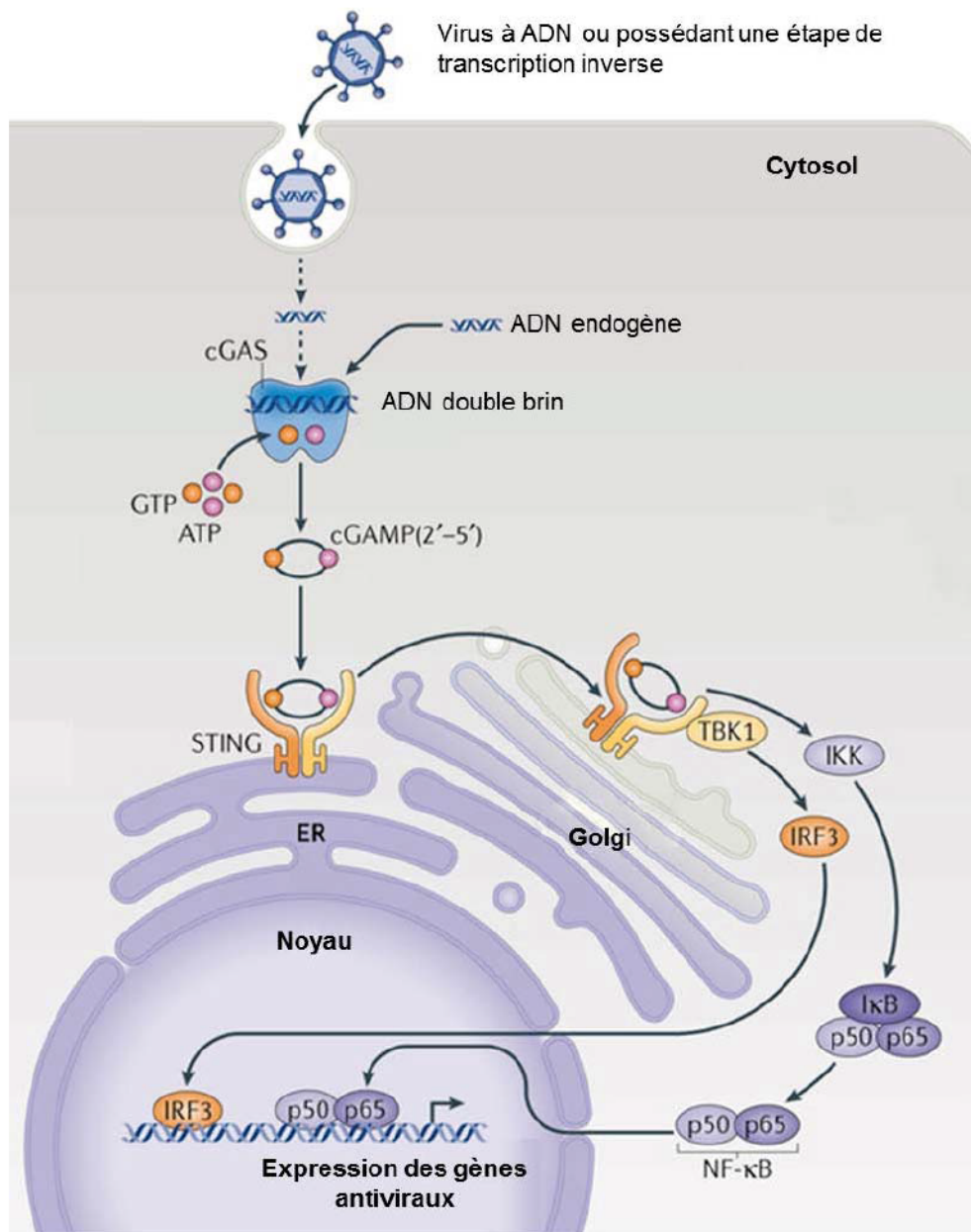


Figure 19 : Mécanisme d'induction de la réponse immunitaire par l'intermédiaire de cGAS et STING. Adapté de Hornung et al, 2014

que facteur de transcription pour les IFN-1 (Figure 18) (Huang & Yang, 2010; J. Wu & Chen, 2014).

Cette production d'IFN-1 peut éventuellement entraîner la surexpression d'A3 par action paracrine. Néanmoins, aucun lien entre l'activation des TLR-9 et la surexpression d'A3 n'a été mis en évidence à ce jour.

b. La voie c-GAS (Cyclic GMP-AMP synthase)

Plus récemment, il a été montré que la protéine Cyclic GMP-AMP synthase (c-GAS) est aussi impliquée dans la détection de l'ADN mitochondrial cytoplasmique. Les auteurs ont montré que ceci aboutit à la mise en place d'une réponse interféron et à l'expression des ISG à un niveau permettant l'inhibition de l'infection par le HSV.

c-GAS appartient à la famille des nucléotidyltransférases et est proche de l'oligoadenylate synthase (OAS). Son domaine catalytique, situé dans la région C-terminal de la protéine, est très conservé parmi les vertébrés. c-GAS reconnaît l'ADN double brin : l'attachement de l'ADN au récepteur va entraîner un changement de conformation du site catalytique : celui-ci s'ouvre et devient accessible à l'ATP et au GTP. La transformation de ces deux molécules en cGAMP se déroule dans une réaction en deux étapes : formation de pppGpA suivie de la cyclisation de celle-ci en cGAMP (Figure 19).

Il a été montré que des cellules telles que les macrophages, les fibroblastes ou les cellules dendritiques *knock out* pour c-GAS ne produisent plus d'IFN1 en réponse à la présence d'ADN double brin cytosolique, démontrant un rôle central de c-GAS dans la mise en place de la réponse immune contre l'ADN cytosolique. Ceci est confirmé par la susceptibilité accrue au HSV-1 des souris *Knock out* pour c-GAS (J. Wu & Chen, 2014).

c. La protéine STING

Le cGAMP produit va ensuite activer la protéine STING (aussi décrite sous les noms de MITA, MPYS, ERIS et TMEM173). Il s'agit d'une protéine cytoplasmique à 4 domaines

transmembranaires localisée à la surface du réticulum endoplasmique. Un cinquième domaine, considéré pendant longtemps comme domaine transmembranaire, est impliqué dans la dimérisation du récepteur. Il s'agit d'une protéine exprimée dans de nombreux types cellulaires. L'activation de STING entraîne une relocalisation de celle-ci au niveau de l'appareil de Golgi. STING va ensuite former des complexes impliquant la kinase TBK-1 et activer celle-ci. Cette kinase phosphoryle le facteur de transcription IRF3 et permet ainsi l'expression des IFN-I (Figure 19).

La protéine STING peut aussi reconnaître directement les dinucléotides cycliques tels que le (3'-5') diguanylate cyclique(c-di-GMP) et le (3'-5') diadenylate cyclique (c-di-ATP) qui sont des messagers intermédiaires dans les voies de régulation bactériennes impliquées dans la virulence, la formation de biofilm ou encore le maintien de l'intégrité génomique.

De la même façon que c-GAS, l'inactivation de STING chez des souris entraîne une sensibilité accrue au HSV-1, ce qui témoigne encore une fois du rôle central de STING dans la réponse immunitaire contre l'ADN cytosolique. De récentes études suggèrent par ailleurs que STING pourrait directement agir en tant que censeur de l'ADN cytosolique. Néanmoins, l'expression de STING dans des cellules déficientes pour cette protéine ne suffit pas à rétablir une réponse interféron lors de stimulation par de l'ADN double brin. De plus, STING possède une affinité 3000 fois moins importante pour l'ADN cytoplasmique que c-GAS ce qui suggère qu'il ne s'agit pas de la fonction principale pour STING (Atianand & Fitzgerald, 2013; J. Wu & Chen, 2014).

Il a été montré que la voie cGAS/STING peut, *in vitro*, détecter l'ADN issu de la rétrotranscription du génome du Virus de la Leucémie Murine (MLV), un rétrovirus murin, et joue un rôle important dans la surexpression des A3. Néanmoins, aucun lien entre la détection de l'ADN mitochondrial cytosolique par cette voie et l'expression des A3 n'a encore été mis en évidence à ce jour (Stavrou, Blouch, Kotla, Bass, & Ross, 2015).

Par ailleurs, il existe d'autres voies impliquées dans la détection de l'ADN cytosolique. Parmi celles-ci, la voie ARN polymérase III/RIG-I.

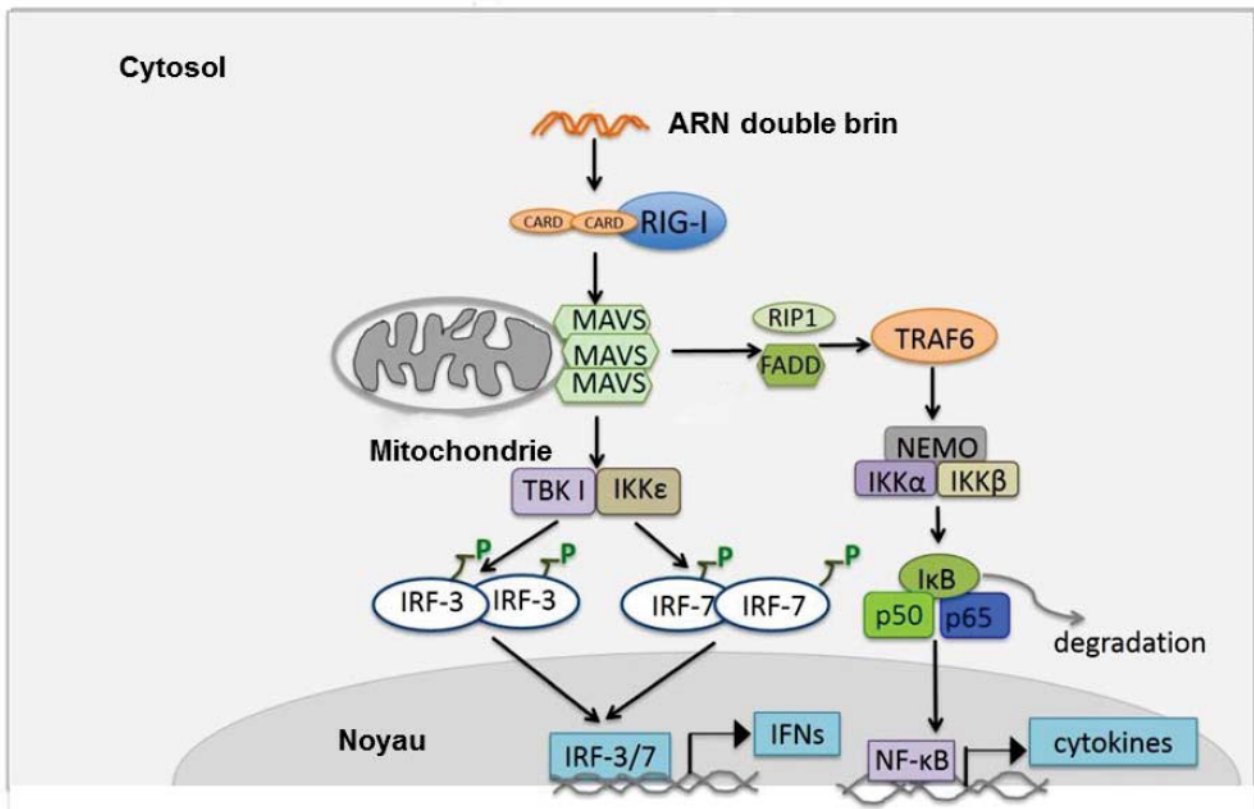


Figure 20 : Mécanisme d'induction de la réponse immunitaire par l'ARN double brin par l'intermédiaire de RIG-I. Adapté de Nan et al, 2014.

d. La voie de RIG-I et de l'ARN polymérase III

RIG-I est une protéine appartenant à la famille des RIG-I-like receptor (RLR) à l'instar de MDA5 et LGP2. Ceux-ci sont caractérisés par leur domaine très conservé hélicase DExD/H-box. Ces RLR reconnaissent l'ARN double brin et jouent donc un rôle clé dans la réponse antivirale contre les virus à ARN double brin tel que les *paramyxoviridae* et les *flaviviridae*.

Néanmoins, il a été montré que RIG-I peut aussi induire une réponse immune en détectant l'ADN double brin poly(dA:dT). Ce paradoxe entre la spécificité de reconnaissance de RIG-I pour l'ARN double brin et la sensibilité de la voie RIG-I à l'ADN double brin poly(dA : dT) a été résolue par la mise en évidence du rôle de la RNA polymérase III dans ce contexte particulier. Celle-ci est principalement impliquée dans la production, à partir de matrices ADN, de petits ARN tels que l'ARN ribosomal 5S, les ARN de transfert ou encore certains micro-ARN. Il a été montré que l'ARN polymérase III encode les ARN viraux de l'EBV (les EBERs 1 et 2) et permet ainsi une réponse interféron par l'intermédiaire de RIG-I.

Une fois RIG-I activée, un changement de conformation du récepteur va permettre la libération des domaines CARD et leur polyubiquitination K63 par les protéines TRIM25 et Riplet. De cette façon, RIG-I va interagir avec la protéine MAVS (aussi appelée IPS-1 ou CARDIF). MAVS est localisée sur la membrane externe de la mitochondrie, à laquelle elle est ancrée par son domaine transmembranaire. Ceci va permettre l'oligomérisation de MAVS et le recrutement de TRAF2, 5 et 6. Ces protéines vont promouvoir des réactions d'ubiquitination afin de recruter la protéine NEMO, qui agit en tant que sous unité régulatrice des kinases IKK et TBK1. Ceci permet de les activer et ainsi activer les facteurs de transcription NFκB et IRF3 (Figure 20) (Atianand & Fitzgerald, 2013; J. Wu & Chen, 2014).

Cette voie permet donc la production d'IFN-I et de cytokines pro-inflammatoires à partir d'ARN double brin viral ou issu de la transcription d'un ADN double brin riche en AT par l'ARN polymérase III.

De façon intéressante, il a été montré que l'activation de RIG-I par de l'ARN double brin 5'triphosphate dans des macrophages infectés par le VIH-1 entraîne une forte production d'IFN-I. Ceci s'accompagne d'une forte inhibition de la réplication virale et d'une surexpression de nombreux facteurs de restriction, dont les A3 (Y. Wang, Wang, Li, Zhou, &

Ho, 2013). Ceci laisse supposer qu'un ADN transcrit par l'ARN polymérase III peut aussi, dans ce type de cellules, entraîner une forte production d'A3.

Qu'en est-il de l'ADN du soi, notamment de l'ADN mitochondrial ?

La présence d'ADN mitochondrial hyperéredité dans le cytoplasme indique qu'un mécanisme permettant l'expression des A3 pourrait être déclenché par la présence d'ADN mitochondrial dans le cytoplasme. Un tel mécanisme nécessite, par ailleurs, la dégradation de l'ADN cytosolique afin d'éviter à la cellule d'être en état d'inflammation chronique

Ces deux problématiques sont à la base de nos travaux présentés dans les résultats (Résultat 3).

Le phénomène d'hypermutation a aussi été observé sur l'ADN nucléaire. En effet, il a été montré par 3DPCR qu'A3A et A3B peuvent induire des mutations sur les gènes *MYC* et *TP53*. Néanmoins, ces mutations sont observables seulement lorsque les cellules sont déficientes pour l'UNG ou en utilisant un inhibiteur de l'UNG : l'Uracil-DNA Glycosylase Inhibitor (UGI) (Stenglein et al., 2010).

3. L'Uracile-ADN glycosylase

L'UNG est une protéine impliquée dans le mécanisme de réparation par excision de bases (BER). Il s'agit du mécanisme permettant de corriger les dommages de l'ADN causés par le métabolisme cellulaire tels que la désamination, l'hydroxylation et la méthylation. La réparation de ces bases va nécessiter l'intervention d'une DNA glycosylase. Ces protéines catalysent le clivage de la liaison N-glycosidique de la base azotée endommagée, la détachant ainsi de l'ADN tout en gardant le squelette sucre phosphate intacte. Ceci crée un site apurique/apyrimidique (AP). Il existe de nombreuses DNA glycosylases parmi lesquelles NEIL 1 à 3, SMUG1, UNG, OGG1, etc. Ces différentes DNA glycosylases vont avoir un substrat différent : par exemple, OGG1 reconnaît les bases 8-oxoguanines tandis que l'UNG reconnaît les uraciles. Etant donné que les A3 induisent des mutations C->U, l'UNG joue un rôle central dans les mécanismes d'*editing* et de réparation.

Il existe deux formes de l'UNG : UNG1 et UNG2. Ces deux formes sont codées par le même gène, *UNG*, mais leurs ARNm respectifs sont issus de deux promoteurs différents : P1

et P2. Ces deux formes diffèrent par leur localisation et leur expression : UNG1 est localisée dans la mitochondrie et est exprimée de manière ubiquitaire tandis que UNG2 est strictement nucléaire et est exprimée dans le placenta, les testicules, le thymus et plus généralement dans les cellules prolifératives. A la suite du clivage par l'UNG, l'AP endonucléase 1 (APE1) va créer un « *nick* » en cassant la liaison phosphodiester et générant ainsi une cassure de l'ADN simple brin formé d'une extrémité 3' hydroxyl (3'OH) et d'une extrémité 5' desoxyribose-5 phosphate (5'dRP) libre. Une base, complémentaire à celle du brin non muté, va être insérée par l'ADN polymérase beta. C'est par ce mécanisme que la grande majorité des mutations par les A3 sont corrigées dans l'ADN génomique. Néanmoins, comme tous les processus biologiques, ce système n'est pas efficace à 100%. C'est de cette façon que quelques uraciles peuvent se maintenir dans l'ADN et engendrer des mutations G->A et C->T.

Ces mutations sont assez rares pour ne pas empêcher la cellule de se diviser et pour se fixer dans le génome et sont qualifiées d'hypomutations.

4. Les hypomutations

Le terme hypomutation a pour la première fois été employé en opposition aux hypermutations somatiques pour des patients souffrants de d'hypogammaglobulinémie commune d'expression variable. Ce syndrome est en réalité un groupe de maladies hétérogènes et aboutissant à un déficit immunitaire qui se caractérise par un taux d'immunoglobuline sériques faible. Parmi ces patients, un certain nombre possèdent dans leur mémoire B des gènes V très peu mutés entraînant une faible diversité d'anticorps (Bonhomme et al., 2000).

La question des hypomutations a ensuite été abordée dans le contexte des mutations du VIH. Il s'agit du cas où les niveaux de mutations induites sur le génome du virus sont compatibles avec la réplication et la transcription ; ainsi, les mutations n'ayant pas d'effet délétère sur le cycle viral sont fixées dans le génome du virus. Ceci peut parfois avoir un effet bénéfique pour le virus puisque ce mécanisme contribue à la diversité génotypique : dans certains cas, ceci permet l'émergence de mutations sur les antigènes viraux et l'échappement au système immunitaire.

De la même façon, on peut se demander si l'ADN génomique peut être le siège d'hypomutations du fait des mutations induites par les A3 nucléaires et la correction imparfaite du BER. Ce mécanisme de mutations non létales pour la cellule permet de fixer des mutations génomiques au fur et à mesure des cycles cellulaires. Ceci pose la problématique de l'impact qu'il pourrait y avoir sur l'émergence de mutations pro-oncogènes et quel rôle les cytidines désaminases peuvent avoir dans la tumorigénèse.

B. La tumorigénèse

L'émergence des cancers est souvent multifactorielle, néanmoins, elle implique souvent une instabilité génomique. Cette instabilité génomique se traduit par des substitutions, des insertions, des délétions ou des recombinaisons résultant de cassures de l'ADN. On distingue deux types de mutations : les mutations dites « *driver* », qui vont conférer un avantage sélectif à la cellule et permettre une multiplication rapide et incontrôlée aboutissant à la formation de tumeurs. D'autres mutations, dites « *passengers* », ne confèrent pas d'avantage particulier à la cellule et n'ont pas d'impact sur l'oncogénèse. Les mutations drivers touchent souvent des gènes impliqués dans le contrôle de la croissance et de la réplication cellulaire. On retrouve mutés les gènes impliqués dans les mécanismes d'apoptose ou dans le cycle cellulaire et la réparation de l'ADN.

Grâce aux progrès du séquençage, il est maintenant possible d'obtenir la séquence complète de différents cancers. Ceci a notamment permis de mettre en évidence la fréquence et la nature des mutations génomiques. Il est intéressant de noter que la fréquence des mutations peut être très variable selon la nature des cancers : en effet, on observe que les cancers pulmonaires et gastriques portent respectivement 4,12 et 2,10 mutations par Mb (Ding et al., 2008; Greenman et al., 2007; Pleasance et al., 2010). En revanche, les cancers des testicules, du sein ou les leucémies portent 0 à 0,19 mutations par Mb (Greenman et al., 2007). Cette forte variabilité dans le nombre de mutations peut s'expliquer en partie par l'origine des cancers et la physiologie des cellules impliquées. Ceci a amené les chercheurs à se pencher sur la nature de ces mutations et d'essayer ainsi de classer ces cancers selon la nature de leurs mutations et ainsi remonter jusqu'à la source des mutations.

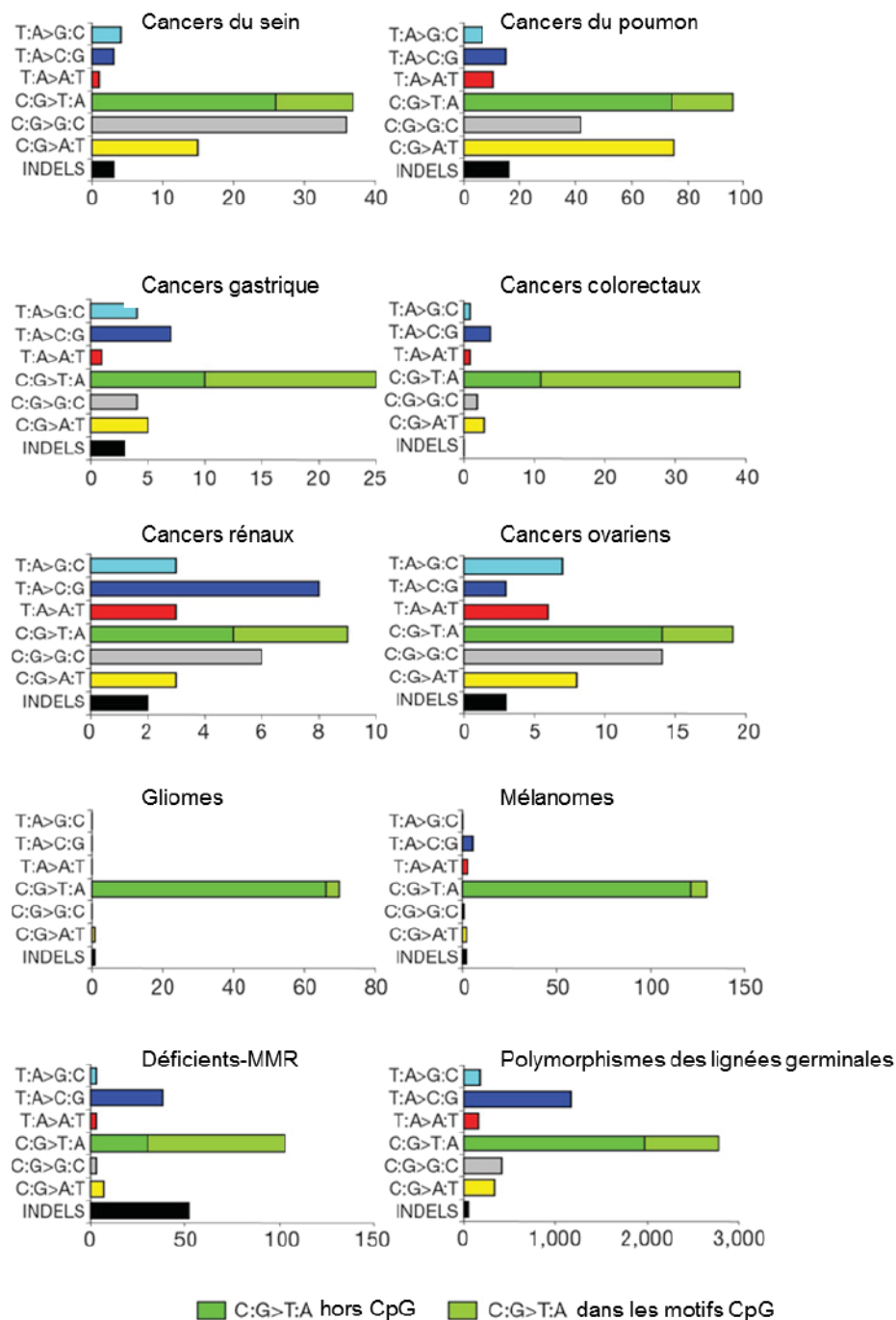


Figure 21 : Spectre des mutations dans les cancers humaines. Les différentes mutations possibles et les indels ont été recherchés. Les mutations CG->AT ont été divisées en deux groupes : hors CpG et dans les ilots CpG. Les polymorphismes sur les lignées germinales sont le résultat du screening de la protéine kinase. Les mutations pour les déficients MMR sont issues de deux cancers colorectaux, deux cancers gastriques et ovariens déficients pour le système de réparation MMR. D'après Greenman et al. 2007.

1. Nature des mutations dans les cancers

Bien que les gènes mutés dans les cancers soient mieux connus, l'origine des mutations reste souvent à déterminer. C'est pour répondre à cette question que de nombreuses équipes se sont intéressées au séquençage de gènes mutés dans les cancers ou de génomes complets dans les cancers. Le séquençage du gène suppresseur de tumeur *TP53* de cellules cancéreuses montre que 55% des mutations sont de type C->T ou G->A (Cheynier, Gratton, Vartanian, Meyerhans, & Wain-Hobson, 1997). Par séquençage complet, on observe que dans les cancers pulmonaires, les cancers gastriques, les cancers ovariens, les gliomes et les mélanomes, la majorité des mutations sont aussi de type G->A ou C->T (Figure 21) (Greenman et al., 2007). Ceci est compatible avec les mutations induites par les A3 mais ne permet néanmoins pas de prouver que les A3 sont responsables de ces mutations. En effet, celles-ci peuvent être le résultat de désaminations spontanées ou d'autres processus aboutissant à l'apparition de ces types de mutations. Seul le contexte dinucléotidique permet d'imputer les mutations aux cytidines désaminases.

En 2012, l'équipe de Michael R. Stratton s'est intéressée à la nature mais aussi au contexte des mutations dans les cancers du sein. Pour cela, ils ont étudié, à partir de 100 cancers du sein, la variation du nombre de copies ainsi que les mutations somatiques de 21 416 gènes et 1664 mirco ARN. Parmi ces 100 échantillons, 79 sont des cancers positifs pour le récepteur aux œstrogènes. Certains génomes sont marqués aussi bien par des transversions que des substitutions tandis que d'autres exhibent un fort biais pour les transitions CG->TA. Ce dernier phénomène est illustré par les auteurs par l'étude des mutations sur le cancer du sein PD4120, positif pour le récepteur aux œstrogènes. Pour cela, ils ont analysé les mutations et leur contexte sur les 21 416 séquences : la majorité des mutations sont, encore une fois, de type CG->TA (Figure 21A). De façon intéressante, le contexte dinucléotidique de ces mutations est très majoritairement le contexte 5'TpC (Figure 21B). Par ailleurs, ces mutations touchent aussi les sites CpG (Stephens et al., 2012).

Dans l'étude suivante, les auteurs ont choisi de réaliser le séquençage complet du génome de 21 cancers du sein. Parmi ces échantillons, on retrouve des cancers positifs pour les récepteurs aux œstrogènes, à la progestérone, pour HER2, un cancer muté dans le gène de prédisposition *BRCA1* et 4 mutés sur le gène de prédisposition *BRCA2*. Dans tous ces cancers, les positions XpCpG sont très fréquemment mutées ainsi que les cytidines méthylées. De plus,

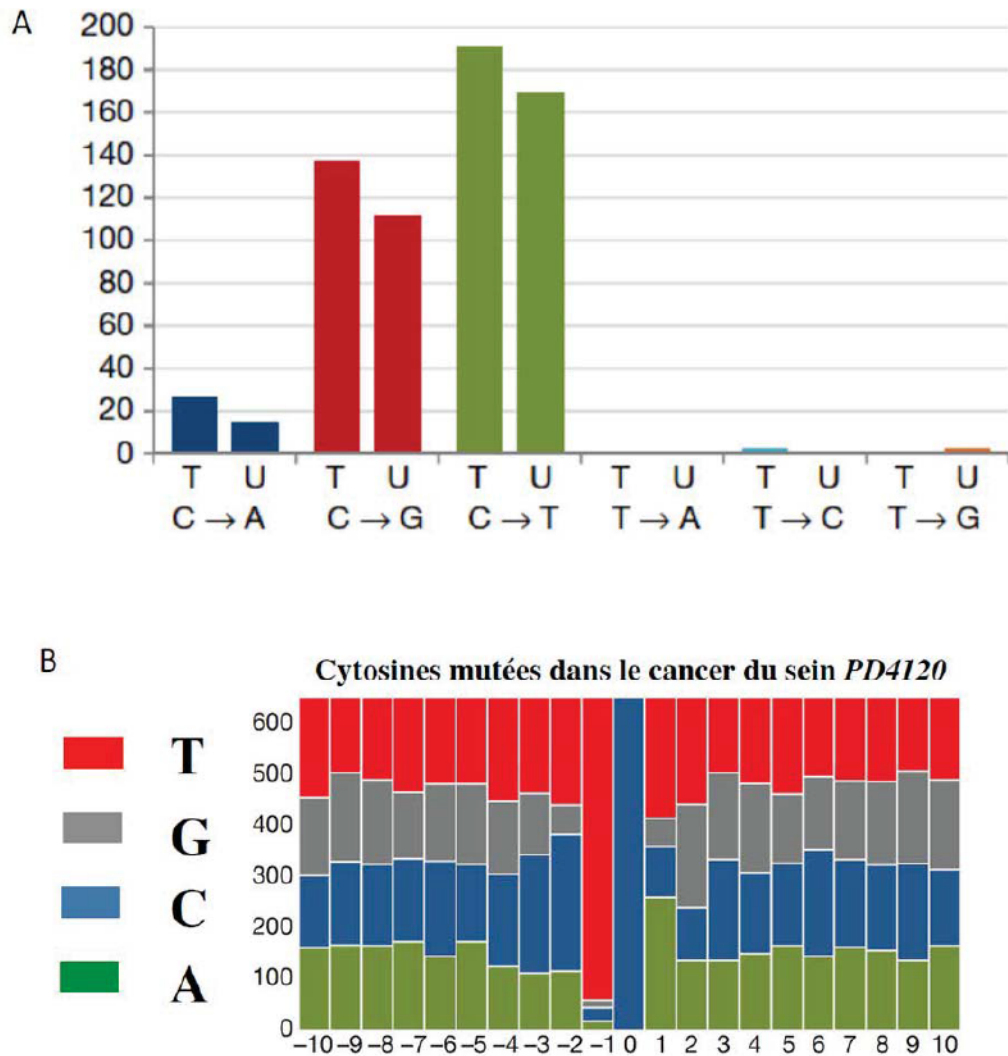


Figure 22 : Nature et signature des mutations C->T dans le cancer du sein PD4120. A. Proportion des mutations dans le cancer du sein PD4120. B. Proportion des bases retrouvées 10 nucléotides en amont (5') et en aval (3') des cytosines mutées dans le cancer du sein PD4120. On observe un fort biais pour le contexte dinucléotidique 5'TpC. D'après Stephen et al, 2012.

la signature TpCpX est très souvent retrouvée. L'analyse plus fine des mutations montre d'une part que les mutations C->T de signature 5'TpC sont situées dans des clusters dans lesquels les mutations sont extensives et monotones. D'autre part, on retrouve dans ces clusters des réarrangements génomiques suivis de mutations de type G->A : celles-ci correspondent aux mutations que l'on retrouverait sur le brin complémentaire. Enfin, l'étude d'un panel plus large de cancers (analyse de 4 938 362 mutations issues de 7042 cancers) montre que 16,6% des mutations se trouvent dans un contexte dinucléotidique compatible avec celui des A3. De plus, des clusters riches en mutations TpCpX et en recombinaisons sont retrouvés aussi bien dans des cancers du sein que des cancers du pancréas, des poumons, du foie, du cervix (médulloblastome), dans des leucémies lymphoïdes chroniques, des leucémies aiguës lymphoblastiques et des lymphomes des cellules B (Nik-Zainal et al., 2012) (Alexandrov, Nik-Zainal, Wedge, Aparicio, et al., 2013).

Ces différentes études mettent en évidence un mécanisme actif entraînant des mutations C->T dans le contexte 5'TpC, mais aussi des recombinaisons consécutives à des cassures de l'ADN double brin. Les membres de la famille APOBEC/AID, et plus particulièrement les A3 nucléaire A3A et A3B, constituent deux candidats logiques dans la survenue de tels événements et l'émergence des cancers. Néanmoins, aucune corrélation directe n'est établie ici. Nous allons donc nous intéresser à l'impact des APOBEC/AID dans la tumorigénèse.

2. Implication d'AID dans la tumorigénèse

La protéine AID, exprimée dans les centres germinaux B, induit des hypermutations somatiques sur les gènes codant les immunoglobulines. Ceci permet la diversification et la commutation de classe de celles-ci. Néanmoins, les mutations induites par AID ne sont pas spécifiques à ce locus. En effet, il a été montré par séquençage haut débit sur des gènes impliqués dans la tumorigénèse des cellules B (*Myc*, *Pim*, *Pax5*, *H2afx*, *Rhoh* et *Ebf1*) issus de cellules B murines que ceux-ci sont sensibles à l'activité d'AID. Les systèmes de réparation de l'ADN permettent ici l'accumulation des mutations. : les auteurs estiment que 25% des gènes exprimés et étudiés ne sont pas intégralement corrigés et portent des mutations. Le taux de mutation est donc le résultat d'un équilibre entre l'activité d'AID et l'efficacité des mécanismes de réparation.

D'autres études montrent que les gènes *Myc*, *Pim*, *Pax5*, *Bcl16*, *CD95* et *Rhoh* portent, dans les cellules B tumorales, des mutations GC->AT, caractéristiques de l'activité d'AID (Muschen et al., 2000; A. Pasqualucci, 1998; L. Pasqualucci, Neri, Baldini, Dalla-Favera, & Migliazza, 2000; L. Pasqualucci et al., 2001). Il a aussi été montré qu'AID était nécessaire à la translocation du gène *MYC* dans le gène *IGH*, caractéristique des lymphomes de Burkitt, ou de la translocation de *BCL2* dans *IGH*, trouvée dans les lymphomes folliculaires et diffus (Jankovic et al., 2010). Ceci soutient l'hypothèse qu'AID peut jouer un rôle pro-tumoral dans les cellules B.

De façon surprenante, AID est aussi exprimée dans des tumeurs solides. Des études sur le cancer du poumon montrent que le taux de mutation sur *TP53* est proportionnel à l'expression d'AID (Shinmura et al., 2011). Il a aussi été observé qu'AID est spécifiquement exprimée dans les cellules de carcinome oral précoce (Miyazaki et al., 2013). AID est aussi responsable de microdélétions dans les cellules infectées par *Helicobacter pylori* (Shimizu et al., 2014).

Ces observations laissent penser que les contextes d'inflammation chronique vont être favorables au développement de tumeurs. Ceci peut s'expliquer par une activation du facteur de transcription NFκB aboutissant à une expression aberrante d'AID. L'expression d'AID a en effet été observée dans des cancers sous inflammation chronique tels que les hépatites virales chroniques.

Par ailleurs, des souris transgéniques possédant le gène *AICDA* sous contrôle d'un promoteur ubiquitaire ont permis d'étudier l'impact d'AID sur la tumorigénèse. Ces souris développent spontanément des lymphomes des cellules T. Les gènes codant les récepteurs des cellules T (TCR), les récepteurs CD4 et CD5, ainsi que les gènes *MYC* et *Pim1* codant des proto-oncogènes, sont très fortement mutés. Néanmoins, aucune translocation de gènes ni de lymphomes B n'est observée. En revanche certaines souris développent des mini-adénocarcinomes des poumons, des carcinomes hépatocellulaires, des mélanomes et des sarcomes (Okazaki et al., 2003). L'activité d'AID est donc variable selon le type cellulaire dans ce modèle et seule l'expression ectopique d'AID mène à la tumorigénèse.

3. Implication de A1 et de A2 dans la tumorigénèse

La protéine A1 chez l'Homme mute spécifiquement l'ARNm de l'ApoB de façon contrôlée. Néanmoins, chez les souris et les lapins transgéniques, cet ARN est muté de façon extensive. Ceci a pour effet de perturber le métabolisme des lipides et aboutit à des dysplasies du foie. Certaines souris développent des carcinomes hépatocellulaires et des hyperplasies du foie. De plus, les ARNm portant des séquences reconnues par A1 sont mutés : ceci suggère qu'A1 peut contribuer à la tumorigénèse en mutant l'ARNm de certaines protéines impliquées dans le cancer (Yamanaka et al., 1995).

Bien qu'A2 ne soit pas active, celle-ci a aussi fait l'objet d'étude sur son éventuelle implication dans la survenue des cancers. L'expression d'A2 de façon ubiquitaire et constitutive chez la souris entraîne l'apparition de tumeurs pulmonaires et de cancers hépatocellulaires. Le taux de mutations sur les transcrits des gènes *Eif4g2* et *PTEN*, souvent mutés dans les cancers, est plus élevé chez les souris transgéniques (Okuyama et al., 2012). Néanmoins, aucun effet d'A2 sur les mutations génomique n'a pu être observé.

4. Implication des A3 dans le cancer

a. Le Rôle d'A3B dans l'émergence des cancers.

L'activité de certaines A3 sur l'ADN génomique et la présence de mutations C->T dans le contexte 5'TpC correspondant aux A3 dans certains cancers ont amené plusieurs équipes à étudier les A3 qui pourraient être impliquées dans le cancer.

A la suite de la découverte des mutations de type A3 dans les cancers du sein, Burns et ses collaborateurs ont recherché dans 38 lignées cellulaires de cancer du sein les A3 pouvant être impliquées. Ils ont pu observer qu'A3B, dont la localisation est strictement nucléaire, est surexprimée dans 28 lignées avec une augmentation pouvant être dix fois supérieure au contrôle. Ils montrent par ailleurs que le taux de mutation dans ces lignées est dépendant d'A3B. Par 3DPCR, les auteurs montrent que le taux de mutation sur *TP53* et *MYC* est plus important dans les lignées exprimant fortement A3B que dans les lignées exprimant peu

d'A3B. Néanmoins, la différence de température de dénaturation reste très faible entre les deux (0,6 °C), ceci se traduisant par des séquences hypomutées (la majorité des séquences ne possède qu'une ou deux mutations C->T par kb). Les auteurs montrent aussi qu'A3B induit des DSB ainsi que des anomalies dans la mitose (cellules anucléées ou multinucléées) et parfois entraîne la mort de la cellule. Ceci n'aboutit pas, contrairement à ce qui a été observé avec A3A, à un blocage du cycle cellulaire en G1/S. Ces observations sont compatibles avec une activité oncogène d'A3B : celle-ci n'est pas assez cytotoxique pour empêcher la division cellulaire mais entraîne des hypomutations et des cassures de l'ADN. Néanmoins, le contexte dinucléotidique n'est pas étudié ici. Par ailleurs, on peut se demander quelle est la pertinence de l'emploi de la 3DPCR alors que celle-ci est initialement prévue pour détecter des hypermutants et non pas le cas particulier des hypomutants (Burns, Lackey, et al., 2013). Néanmoins, l'analyse de données de séquençage issues de trois cancers du sein différents montre une fréquence de mutations C->T de 40% avec comme contexte dinucléotidique majoritaire le contexte TpC, et un taux d'A3B élevé.

A3B est retrouvée surexprimée dans d'autres cancers : cancers du col de l'utérus, cancer de la vessie, cancer des voies aérodigestives supérieures et cancer des poumons (Burns, Temiz, et al., 2013). Dans ces cancers, de nombreuses mutations C->T sont retrouvées mais le contexte TpC ne se démarque vraiment que dans les cancers de la vessie et du col de l'utérus. D'un point de vue mécanistique, la surexpression d'A3B pourrait être dans certains cas induite par une infection virale : ceci est observé dans certains cas de cancers du sein positifs pour les récepteurs aux œstrogènes. Dans cette étude, il a été montré que les protéines E6 et E7 de HPV 18 induisent la surexpression d'A3B. Enfin, il a aussi été montré que des hauts taux d'A3B sont corrélés avec des pronostics défavorables dans des cancers du sein positifs pour les récepteurs des œstrogènes.

b. Impact de la délétion d'A3B sur la carcinogénèse

Une étude *in vitro* chez la levure confirme le rôle que pourrait avoir A3B dans les mutations à l'origine des cancers. Néanmoins, cette étude montre aussi qu'A3A peut être impliquée dans l'apparition de cancers et que son activité est indiscernable de celle d'A3B étant donné que les deux éditent dans le même contexte dinucléotidique. De plus, une étude

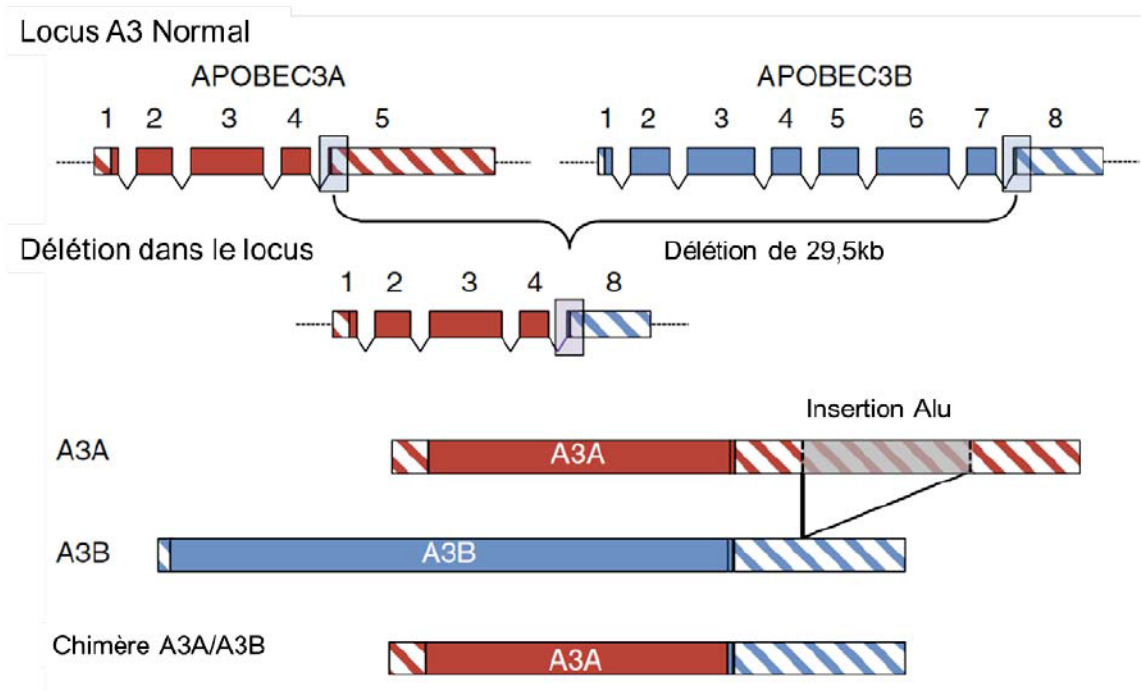


Figure 23 : Schéma du locus A3 normal ou $\Delta A3B$ portant une délétion de 29.5kb. Les rectangles hachurés représentent les régions 5' et 3' non transcrites. La couleur rouge est associée aux séquences du gène *A3A* tandis que la couleur bleue est associée aux séquences issues du gène *A3B*. Les chiffres représentent les numéros des exons. Un rectangle opaque représente l'insertion de la séquence Alu présente dans le 5'UTR d'*A3A* et responsable de la faible stabilité de son ARNm. D'après Caval et al. 2014.

japonaise datant de 2008 montre qu'une délétion d'A3B est retrouvée sous forme homozygote dans 16,7% des tumeurs étudiées et sous forme hétérozygote dans 30% des tumeurs étudiées. Cette délétion est presque fixée dans la population Océanienne (92%) et très commune dans l'est de l'Asie et les populations amérindiennes (36,9% et 57,7% respectivement). En revanche, elle est plutôt rare dans les populations caucasiennes et africaines (Kidd et al., 2007). Cette délétion est aussi associée à une plus grande susceptibilité à l'infection par *Plasmodium falciparum* (Jha et al., 2012), à l'insertion d'un élément Alu dans le gène de l'Angiotensine Converting Enzyme (K. Wang et al., 2013) mais aussi à l'infection au VIH-1 (An et al., 2009), bien que ce point reste encore controversé.

Dans le cadre du cancer, cette délétion est également associée à une plus grande susceptibilité à la persistance du VHB et au développement de carcinomes hépatocellulaires mais aussi au développement de carcinomes épithéliaux de l'ovaire et au cancer du pancréas (T. Zhang et al., 2012). Enfin, une autre étude montre que le risque de cancer du sein chez les femmes d'ascendance européenne est augmenté de 2 fois en cas de délétion homozygote d'A3B et que le nombre de mutation est plus important dans ce cas (Xuan et al., 2013).

Comment le risque de cancer peut-il être augmenté et le nombre de mutations plus important en l'absence de la protéine A3B alors qu'il a été montré que celle-ci contribuait aux mutations à l'origine des cancers ?

La délétion en question est longue de 29,5kb et couvre une région allant de l'exon 5 d'A3A à l'exon 8 d'A3B. Il en résulte une séquence composée des 4 premiers exons composant A3A, l'exon 8 d'A3B et de la partie 3' Non traduite (UTR) d'A3B. Le gène ainsi formé, une fois transcrit, permet la production d'un ARNm ne possédant non plus l'UTR d'A3A mais celui d'A3B (Figure 23).

L'ARNm d'A3A possède une séquence Alu dans son 3'UTR : il a été montré que les ARNm d'A3A sans 3'UTR ou avec la séquence 3'UTR d'A3B sont plus stable que l'ARNm d'A3A sauvage. Ceci se traduit par un taux de DSBs et de mutations plus important lorsque ces ARNm d'A3A sont transfectés. De plus, l'induction d'A3A dans une lignée SKBR3, déficiente pour A3B, entraîne des DSBs et des mutations de l'ADN génomique. Ceci permet donc d'expliquer comment la délétion d'A3B peut aboutir à une plus grande sensibilité au cancer. Par ailleurs, cette étude montre que seule A3A génère des DSBs, et pas A3B comme

montré précédemment et qu'A3A est plus active sur les cibles nucléaires qu'A3B (Caval, Suspene, Shapira, et al., 2014).

Bien qu'A3A soit plus active, celle-ci est très finement régulée à plusieurs niveaux. Au niveau transcriptionnel, on peut observer que les codons d'initiations de toutes les *A3* sont situés dans un « bon contexte » de Kozak, à l'exception d'*A3A*. En effet, les *A3* possèdent une séquence consensus (gcc)gccR₃ccAUG G₊₁ (avec R correspondant à un A ou un G) ce qui n'est pas le cas d'A3A. Ceci se caractérise par un très faible niveau d'expression basal d'A3A. De plus, A3A est exprimée dans un nombre restreint de cellules témoignant d'une régulation particulière de l'expression de cette protéine. L'analyse de la séquence d'*A3A* montre que celle-ci possède deux codons d'initiation codant deux isoformes : A3A_{P1} et A3A_{P2}, plus courte de 12 acides aminés et uniquement présente dans la lignée myéloïde et non dans la lignée T (Thielen et al., 2010). Bien qu'identique d'un point de vue enzymatique, la forme P2 n'est exprimée que si le codon d'initiation de la forme P1 est dans un mauvais contexte de Kozak. Des expériences visant à placer le codon d'initiation de P1 dans un bon contexte montrent que ceci a pour effet d'abolir l'expression de la forme P2. L'expression de la seconde isoforme est donc dépendante de la première. Néanmoins, la fonction de ce mécanisme de régulation est encore inconnue.

A3A est aussi régulée au niveau protéique. En effet, il a été montré que celle-ci interagit et est dégradée par la protéine Tribble3 (TRIB3) (Aynaud et al., 2012). Cette protéine fait partie d'une famille de trois membres (TRIB1 à TRIB3) et est codée par le gène *TRIB3* situé sur le chromosome 20. Les premières études sur TRIB3 montrent que celle-ci interagit avec des protéines impliquées dans l'apoptose : ATF4, CHOP, p53/RelA. TRIB3 joue également un rôle dans la régulation du cycle cellulaire, au niveau des points de contrôle G2/ et G1/S. Ceci implique une interaction avec p53/RelA, cdc25, CtiP et MAPKK (Kiss-Toth et al., 2004; Mayumi-Matsuda, Kojima, Suzuki, & Sakata, 1999; Sakai et al., 2010; J. Xu et al., 2007).

TRIB3 a une localisation strictement nucléaire et possède un domaine pseudo-kinase mais aucune activité kinase n'a pu être mise en évidence. Néanmoins, ce domaine est très conservé dans l'évolution. Il a été montré récemment que TRIB3 interagit avec A3A. Cette interaction a pour effet d'induire la dégradation d'A3A par des voies indépendantes du protéasome et du lysosome. Ceci a pour effet d'inhiber l'*editing* et la formation de DSB induits par A3A. L'étude de l'expression de TRIB3 dans les PBMC par PCR quantitative montre que les

niveaux de TRIB3 dominant ceux d'A3A, ce qui suggère qu'A3A est systématiquement sous le contrôle de TRIB3 (Aynaud et al., 2012). Cette protéine joue donc le rôle de modulateur de l'activité d'A3A et se place en gardien de l'intégrité du génome. Ceci s'intègre parfaitement aux autres fonctions de TRIB3 dans la régulation de l'apoptose et la régulation du cycle cellulaire.

Nous nous sommes par conséquent intéressés aux différences entre A3A et A3B et à la façon dont A3B a divergé d'A3A, menant à une A3 moins active et donc moins dangereuse pour l'intégrité génomiques. Ceci sera traité dans les résultats 3.

Les A3 sont donc impliquées dans des mécanismes de restriction virale mais semblent aussi jouer un rôle important dans la dégradation de l'ADN cytosolique, dans l'*editing* de l'ADN génomique et la carcinogénèse. Ceci peut être la conséquence d'une évolution rapide vers une diversification de cette famille. Afin de protéger la cellule contre ces puissants mutateurs, des éléments de régulation sont apparus au cours de l'évolution. C'est le cas par exemple de la protéine Tribble3 qui inhibe et dégrade la protéine A3, protégeant ainsi le génome contre son activité.

Problématiques et objectifs des travaux de thèse.

Les cytidines désaminases APOBEC/AID constituent des protéines centrales du système immunitaire qu'il soit inné (A3) ou adaptatif (AID). L'importante diversité parmi les A3 chez les primates a permis une diversification des rôles dans la défense antivirale et une réponse adaptée à différents virus. Néanmoins, certaines APOBEC n'ont pas d'activité ni e rôle connu chez l'Homme. C'est le cas d'A3DE, qui porte une mutation, par comparaison avec A3F, en position 320 d'une tyrosine vers une cystéine chez l'Homme et le Gorille. En effet, la réversion de cette mutation permet de restaurer l'activité cytidine désaminase sur le VIH. Malgré cette mutation, A3DE demeure surexprimée dans les tissus cirrhotiques infectés par le VHB. De plus, A3DE peut former des dimères avec A3F et A3G, sans que la fonction de cette interaction soit connue. Nous nous sommes par conséquent demandé quel est le rôle d'A3DE, dans le contexte de l'infection par le VHB. Nous avons pour cela recherché la forme active dans les cirrhoses de patients infectés mais nous avons aussi étudié le rôle de l'interaction d'A3DE avec A3F et A3G (**Résultat 1**).

A3B, contrairement à A3A, est une A3 strictement nucléaire. Il a été montré que ces deux cytidines désaminases induisent des mutations sur l'ADN génomique et que celles-ci pourraient être impliquées dans la genèse des cancers. A3A et A3B sont deux enzymes assez proches puisque A3A est composée d'un domaine Z1 tandis qu'A3B est composée de deux domaines : un domaine N-terminal Z2 et un domaine C-terminal Z1. Le domaine C-terminal étant porteur de l'activité cytidine désaminase des A3, la protéine A3A et A3B partagent le même type de domaine catalytique. Néanmoins, A3B n'induit pas de cassures de l'ADN double brin, contrairement à A3A. Ceci nous a amené à nous demander quelles différences entre ces deux protéines mènent à une atténuation de l'activité d'A3B par rapport à A3A. Par comparaison de séquences et mutagenèse, nous avons pu caractériser les déterminants moléculaires ayant mené A3B à évoluer vers une atténuation de son activité par rapport à A3A (**Résultat 2**).

L'activité atténuée d'A3B et son rôle probable dans la genèse de certains cancers, notamment les cancers du sein ont aussi été pour nous l'occasion de considérer la partie technique de la détection des désamination. En effet, la 3DPCR, méthode mise au point dans notre laboratoire afin d'amplifier spécifiquement par PCR les cibles ADN hypermutés pas les

A3, est employée aussi bien pour l'amplification de séquences virales hypermutées que pour l'amplification de séquences d'ADN génomique hypermutés et hypomutés. Ce dernier cas est problématique car cette technique, basée sur l'amplification à de multiples températures de dénaturation, permet de mettre en évidence des événements d'hypermutations mais ne permet pas de différencier des séquences hypomutées de séquences non mutées. Ceci peut aboutir à une interprétation erronée de résultats, notamment dans l'étude du lien entre A3 et cancers. Nous avons donc étudié quelles sont les limites de la 3DPCR et de quelle façon il est possible, combinée à d'autres techniques, de l'employer pour la détection de séquences hypomutées (**Résultat 3**).

Enfin la mise en évidence dans notre laboratoire de séquences d'ADN mitochondrial cytoplasmique hypermutées par les A3 a orienté nos recherches vers le rôle des A3 dans le catabolisme de l'ADN cytosolique. Nous avons pour cela recherché l'origine de ce phénomène. Nous avons dans un premier temps évalué la fréquence des mutations sur l'ADNcymt *in vivo* sur des lymphocytes T CD4⁺ et sur l'ADNmt circulant de patients atteints de sepsis. Nous avons par la suite étudié dans la lignée cellulaire THP-1, exprimant faiblement A3A au niveau basal mais dont l'expression est fortement augmentée par l'interféron, l'impact de la présence d'ADN cytosolique sur la cellule, sur l'expression des A3, et sur l'*editing* de l'ADNmt et de l'ADN nucléaire (**Résultat 4**).

MATERIELS ET METHODES

I. Résultat 1

A. Patients

Les échantillons prélevés sur des patients atteints de cirrhose proviennent d'une étude précédente (Vartanian et al., 2010). Tous les échantillons provenant de patients atteints de carcinome hépatocellulaire ont été décrits dans des publications précédentes (Marchio et al., 2000; Pineau et al., 2008; Pineau et al., 1999). Le recrutement des patients a été fait en accord avec la déclaration d'Helsinki et a été approuvé par un Comité de Protection des Personnes (Institutional Review Board, RBM n°2005-019). Le consentement de chaque patient a été obtenu.

B. Lignées cellulaires et cultures

Les lignées cellulaires Huh7, HeLa et HEK293T ont été maintenues en culture dans un milieu DMEM complété avec du sérum de veau fœtal à 10%, 100 unités/ml de pénicilline et 100 µg/ml de streptomycine (GIBCO/BRL) à 37°C et 5% de CO₂. La lignée QT6 a été maintenue en culture dans un milieu Ham's complétée avec du sérum de veau fœtal à 10%, 100 unités/ml de pénicilline et 100 µg/ml de streptomycine (GIBCO/BRL), 2mM de glutamine, 5% de tryptose phosphate, 1% de sérum de poulet à 37°C et 5% de CO₂

Environ 450 000 cellules QT6 ont étéensemencées sur des plaques 6 puits et transfectées 24 h après avec 2 µg de plasmide total à l'aide de JetPrime (Polyplus Transfection TM). Après 48 h, l'ADN total a été extrait à l'aide du kit EPICENTRE. Pour la microscopie, 50 000 cellules HeLa ont étéensemencées sur des lames Lab-Tek et transfectées 24 h après avec 1 µg de plasmide total à l'aide de l'agent de transfection FuGENE HD Transfection Reagent (Roche Applied Science). Pour la qPCR, environ 450 000 cellules Huh7 ont étéensemencées sur des plaques 6 puits et transfectées 24h après avec 2 µg de plasmide total avec l'agent de transfection JetPrime (Polyplus Transfection TM). Le milieu a été changé après 24 h. Après 48 h ou sept jours de culture, le surnageant a été traité avec 10 U de DNase I et incubé 2h à 37°C puis inactivé à 65°C. L'ADN a ensuite été extrait du surnageant à l'aide du kit EPICENTRE (Illumina).

C. Plasmides

Les ADNc codant A3A à A3H ont été amplifiés par PCR standard (*Taq* Platinum, Invitrogen) et clonés dans le vecteur pcDNA3.1/V5-His TOPO à l'aide du kit pcDNA3.1/V5-His TOPO Expression Kit (Invitrogen) et par recombinaison *in vitro* dans le plasmide pDONR207 à l'aide du kit Gateway (Invitrogen). Les amorces de PCR portent 20 à 30 nucléotides s'hybridant avec la séquence d'intérêt afin d'avoir un T_m proche de 60°C. L'extrémité des amorces sens pour la recombinaison est fusionnée à la séquence de recombinaison attB1.1 5'GGGGACAACCTTTGTACAAAAAAGTTGGCATG-3'tandis que l'amorce anti-sens est fusionnée à la séquence de recombinaison attB2.1 5'-GGGGACAACCTTTGTACAAGAAAGTTGGTTA-3. Les produits de PCR ont ensuite été clonés dans pDONR207 selon les recommandations du fabricant (BP cloning reaction, Invitrogen). Les constructions ont été transformées et amplifiées dans la souche E. Coli TOP10. A3BnA a été synthétisée avec l'intron 7 de sa séquence et cloné dans pDONR207 de la même façon que les autres séquences. Le siARN d'A3DE est un mélange des séquences 5'GUUUGAGAACCACGCAGAA et 5' UUCUGCGUGGUUCUCAAAC. Le siARN A3DE et le siARN universal negative control (SIC001) proviennent de chez Sigma-Aldrich

D. Extraction protéique et purification de co-affinité

Afin de réaliser les expériences de co-affinité, les séquences codantes contenues dans pDONR207 ont été transférées dans le vecteur d'expression pDEST27 (Invitrogen) afin que les protéines soient fusionnées à la GST et dans le plasmide pCI-neo-3xFLAG afin que les protéines soient fusionnées au *tag* 3xFLAG. 300 ng de construction GST et 100 ng de construction 3xFLAG ont ensuite été transfectées dans la lignée HEK-293T 24h après ensemencement de celles-ci. Deux jours après transfection, les cellules ont été lavées au PBS et resuspendues et dans un tampon de lyse (0.5% Nonidet P-40, 20 mM Tris-HCl, pH 7.4, 120 mM NaCl, et 1 mM EDTA) supplémenté avec un inhibiteur de protéase (Complete Protease Inhibitor, Roche Applied Science). Les lysats cellulaires ont été incubés 20 min sur glace et clarifiés par centrifugation à 14 000g pendant 30 min. Pour le GST pulldown, 400 µg de protéines ont été incubées pendant 1h à 4°C avec 25 µl de billes glutathione-Sepharose (Amersham Biosciences) afin de purifier les protéines fusionnées à la GST. Les billes ont ensuite été lavées 3 fois avec du tampon de lyse froid. Les protéines ont ensuite été récupérées en chauffant le mélange à température de dénaturation dans du tampon de charge (Invitrogen).

E. Western Blot

Les complexes purifiés et les protéines extraites ont été séparés par électrophorèse SDS-PAGE sur un gel 4-12% NuPAGE Bis-Tris et du tampon de migration MOPS (Invitrogen). Les protéines ont ensuite été transférées sur membrane de nitrocellulose. Les protéines fusionnées avec 3xFLAG, V5 ou GST ont été mises en évidence par les techniques d'immunoblot standard. Les anticorps utilisés sont un anticorps monoclonal de souris anti-3xFLAG conjugué à la HRP (M2; Sigma) ou un anticorps de lapin polyclonal anti-GST (Sigma) ou un anticorps monoclonal de souris anti-V5 conjugué à la HRP (Invitrogen). Un anticorps secondaire anti-lapin conjugué à la HRP a été utilisé le cas échéant (Amersham Biosciences).

F. Extraction d'ADN, PCR et 3DPCR.

Les ADN totaux des cellules transfectées ont été extraits à l'aide du kit MasterPure™ complete DNA and RNA purification kit (Epicentre) et resuspendus dans 30 ml d'eau stérile. Les échantillons de tissus non tumoraux ont été digérés dans un tampon SDS/protéinase K à 37°C durant 6 à 12h, puis l'ADN extrait au phénol/chloroforme. L'ADN a ensuite été précipité à l'éthanol et resuspendu dans un tampon TE. Pour l'amplification du gène *X* du VHB, la première étape de PCR a été réalisée de la façon suivante : 95°C durant 5 min suivis de 35 cycles (95°C durant 1 min, 50°C durant 30 sec, 72°C durant 5 min) et enfin 10 min à 72°C avec pour amorces 5'CGCAAATATACATCGTATCCAT et 5'AAGAGTYTYTTATGTAAGAGCYT. Pour la seconde étape de PCR, l'équivalent de 0,5 µg de la première étape a été utilisé et la réaction a été réalisée sur un thermocycleur Eppendorf gradient Mastercycler S. Les paramètres de réaction sont les suivants : 80-92°C durant 5 min puis 35 cycles (80-92°C durant 1 min, 45°C durant 30 sec et 72°C durant 2 min) and finalement 10 min à 72°C. Les amorces utilisées sont 5'ATGGCTGCTARGCTGTGCTGCCAA et 5'AAGTGCACACGGTYTGGCAGAT, où Y=T/C et R=A/G.

Pour l'amplification de l'exon 6 d'A3DE, 25ng d'ADN génomique ont été amplifiés sur 35 cycles en employant un protocole "step-down" dans un thermocycleur Techne Genius

(OSI, Saint-Quentin-en-Yvelines, France). Une première étape d'amplification avec des amorces ancrées dans la séquence spécifique d'A3DE a été utilisée, suivis d'une seconde étape de PCR interne. La première amplification a été réalisée avec les amorces 3DEes, 5'ACCCCAGCTTGGGCAACAGG et 3DEer, 5'TGCAGCTGGCGCCGGCCCCA. Les réactions ont été réalisées dans un mix de 25 µl contenant un tampon de PCR 1X (50 mM KCl, 10 mM Tris pH 8.9, 0.1% Tween 20, 1.5 mM MgCl₂), 200 µM dNTP, 20 pmoles de chaque amorce et 1 U de *Taq* polymérase. Les conditions d'amplification sont 95°C durant 5 min suivis par 35 cycles (95°C durant 30 sec, 52°C durant 30 sec, 72°C durant 1 min) et finalement 10 min à 72°C. La PCR nichée a été réalisée avec les amorces 3DE320p, 5'TGTCTCCTAACACAACTACG et 3DE320j, 5'CCCCTCCTGGTAATCTGTATCCCAGAACTAG. Les produits d'amplification ont été purifiés sur gel d'agarose (Qiaex II kit, Qiagen, France) et ligués dans le vecteur de clonage TOPO TA (Invitrogen, France). Le séquençage a été réalisé par GATC. Toutes les séquences ont été vérifiées sur chromatogramme.

F. PCR quantitative

Les niveaux d'expression des *A3*, de *RPL13A* et de l'ADN du VHB ont été quantifiés par qPCR à l'aide d'un Eppendorf Realplex2. Les réactions ont été réalisées sur une plaque 96 puits contenant pour chacun des puits 12.5 µL de 2x probe Master Mix (Roche), 2 µl d'amorces (5 µM chacune), 0.33 µL d'UPL probe (Roche), 5.17 µL d'eau et 5 µl d'ADN). Le programme employé est le suivant: 95°C 10min suivi de 45 cycles (95°C 15 sec, 58°C 15 sec, 68°C 30 sec). Pour la quantification du niveau d'expression des *A3* et de *RPL13A*, les amorces et sondes employées sont celles décrites par (ref). Les résultats sont normalisés par rapport aux niveaux d'expression de *RPL13A*. Pour la quantification de l'ADN VHB, les amorces employées sont celles de la première étape de 3DPCR sur *X* en combinaison avec la sonde UPL125 (Roche).

G. Microscopie confocale

Après avoir lavé les cellules au PBS, celles-ci ont été fixées à l'aide de paraformaldéhyde 4% durant 20 min à température ambiante. Les cellules ont ensuite été perméabilisées avec un mélange 50/50 de méthanol/acétone durant 20 min à température ambiante. Les anticorps de souris anti-V5 (Invitrogen) et de lapin anti-Flag (Sigma) ont été

utilisés dilués au 1/200 durant 1h à température ambiante. Les cellules ont ensuite été incubées avec des anticorps couplés à l'Alexa 488 ou Alexa 555 (Invitrogen) dilués au 1/200 durant 1 h à température ambiante et à l'obscurité. Les lames ont ensuite été montées avec le milieu de montage Vectashield-DAPI. L'imagerie a été réalisée sur un microscope Leica-TCS-SP5 équipé d'un objectif X1.2 WATER.

L'analyse de colocalisation a été réalisée à l'aide du logiciel Huygens Essential for mac version 4.3.1p3 (Scientific Volume Imaging B.V.). Les images ont d'abord été déconvoluées puis analysées par la méthode de Costes.

Plasmid	Matrix	Primers
A3Bc	A3Bi7	For : 5'-CACC AT GGATCCAGACACATTCACCTTCAACTTTA-3' Rev : 5'-GTTTCCCTGATTCTGGAGAAT-3'
A3BnA	Puc57-A3BnA	For : 5'-CACC AT GAATCCACAGATCAGAAATCCGATGGAGC-3' Rev : 5'-GTTTCCCTGATTCTGGAAGAGCAGGGGGGTG-3'
A3BnA1	A3BnA	For : 5'-CACC AT GAATCCACAGATCAGAAATCCG-3' Rev : 5'-GCCATTGTCCAGGCGCTCCACTTCGTAGCACAG-3'
	A3Bi7	For : 5'-CTGTGCTACGAAGTGGAGCGCCTGGACAATGGC-3' Rev : 5'-GTTTCCCTGATTCTGGAAGAGCAGGGGGGTG-3'
A3BnA2	A3BnA	For : 5'-CACC AT GAATCCACAGATCAGAAATCCG-3' Rev : 5'-GCCATTGTCCAGGCGCTCCACTTCGTAGCACAG-3'
	A3Bi7	For : 5'-CTGTGCTACGAAGTGGAGCGCCTGGACAATGGC-3' Rev : 5'-TATAAGGAGGCGCTGCAAATGCTGCGGGATGCT-3'
A3BnA3	A3Bi7	For : 5'-CACC AT GAATCCACAGATCAGAAATCCG-3' Rev : 5'-AGCATCCCGCAGCATTTGCAGCGCCTCCTTATA-3'
	A3BnA	For : 5'-TATAAGGAGGCGCTGCAAATGCTGCGGGATGCT-3' Rev : 5'-GTTTCCCTGATTCTGGAAGAGCAGGGGGGTG-3'
A3BnA1-2	A3BnA	For : 5'-CACC AT GAATCCACAGATCAGAAATCCG-3' Rev : 5'-AGCATCCCGCAGCATTTGCAGCGCCTCCTTATA-3'
	A3Bi7	For : 5'-TATAAGGAGGCGCTGCAAATGCTGCGGGATGCT-3' Rev : 5'-GTTTCCCTGATTCTGGAAGAGCAGGGGGGTG-3'
A3BnA1-3	A3BnA	For : 5'-CACC AT GAATCCACAGATCAGAAATCCG-3' Rev : 5'-AGCATCCCGCAGCATTTGCAGCGCCTCCTTATA-3'
	A3Bi7	For : 5'-TATAAGGAGGCGCTGCAAATGCTGCGGGATGCT-3' Rev : 5'-GTTTCCCTGATTCTGGAAGAGCAGGGGGGTG-3'
A3BnA2-3	A3Bi7	For : 5'-CACC AT GAATCCACAGATCAGAAATCCG-3' Rev : 5'-GCCATTGTCCAGGCGCTCCACTTCGTAGCACAG-3'
	A3BnA	For : 5'-CTGTGCTACGAAGTGGAGCGCCTGGACAATGGC-3' Rev : 5'-GTTTCCCTGATTCTGGAAGAGCAGGGGGGTG-3'
A3Bc1	A3BnA1	For : 5'-CACC AT GGATCCACACATATTCACCTCC-3'
A3Bc1-2	A3BnA1-2	Rev : 5'-GTTTCCCTGATTCTGGAAGAGCAGGGGGGTG-3'
A3Bc1-3	A3BnA1-3	
A3Bc2	A3BnA2	For : 5'-CACC AT GGATCCACACATATTCACCTCC-3'
A3Bc2-3	A3BnA2-3	Rev : 5'-GTTTCCCTGATTCTGGAAGAGCAGGGGGGTG-3'
A3Bc3	A3BnA3	
Rhesus-A3B	Puc57-RhesusA3B	For : 5'-CACC AT GAATCCACAGATCAGAAATCCG-3' Rev : 5'-GTTTCCCTGATTCTGGAGAATGGC-3'
Rhesus-A3Bc	Puc57-RhesusA3B	For : 5'-CACC AT GGATCCAGACACGTTCACTTCCAA-3' Rev : 5'-GTTTCCCTGATTCTGGAAGAGCAGGGGGGTG-3'
Rhesus-A3Ap2	Rhesus-A3A	For : 5'-CACC AT GGATCCAAACACGTTCACTTTCAAC-3' Rev : 5'-GTTTCCCTGATTCTGGAGAATGGC-3'
A3Ap2HI16DT	A3A	For : 5'-CACC AT GGATCCAGACACATTCACCTT-3' Rev : 5'-GTTTCCCTGATTCTGGAGAATGGC-3'
A3Bcd96H	A3Bc	For : 5'-CACC AT GGATCCACACACATTCACCTT-3' Rev : 5'-GTTTCCCTGATTCTGGAGAAT-3'
A3Bct97I	A3Bc	For : 5'-CACC AT GGATCCAGACATATTCACCTT-3' Rev : 5'-GTTTCCCTGATTCTGGAGAAT-3'

Tableau 2 : Liste des amorces employées pour les constructions des chimères A3A/A3B pour le système pcDNA3.1 TOPO TA V5 cloning.

II. Résultat 2

A. Plasmides

Les plasmides exprimant A3A, A3Ap2 et A3B et leurs mutants catalytiques proviennent d'une étude précédente (Caval, Suspene, Shapira, et al., 2014; Mussil et al., 2013; Suspène, Aynaud, Guétard, et al., 2011) et ont été utilisés pour la construction de chimères par PCR (Tableau 2) ou pour la mutagenèse dirigée (GeneArt® Site-Directed Mutagenesis System, Life Technologies) (Tableau 3). Afin de palier à la toxicité d'A3BnA et d'A3B de rhésus (rhA3B) chez *E. coli*, les plasmides ont été synthétisés avec l'intron 7 et clonés dans le vecteur pcDNA3.1D/V5-His-TOPO (Life Technologies). Le plasmide exprimant l'A3A de rhésus est issu d'une étude antérieure et a été utilisé pour générer le plasmide Rhesus-A3Ap2 (ref) (Caval, Suspene, Vartanian, et al., 2014). Toutes les constructions ont été clonées et amplifiées dans la souche TOP10 d'*E. coli* (Life Technologies) et vérifiées par séquençage.

B. Culture cellulaire

La lignée QT6 a été maintenue en culture dans un milieu Ham's complétée avec du sérum de veau fœtal à 10%, 100 unités/ml de pénicilline et 100 µg/ml de streptomycine (GIBCO/BRL), 2mM de glutamine, 5% de tryptose phosphate, 1% de sérum de poulet à 37°C et 5% de CO₂. Les lignées HeLa et 293T-UGI, cellules exprimant l'inhibiteur de l'UNG (UGI) du phage *Bacillus subtilis*, ont été maintenues en culture dans un milieu DMEM complété avec du sérum de veau fœtal à 10%, 100 unités/ml de pénicilline et 100 µg/ml de streptomycine (GIBCO/BRL) à 37°C et 5% de CO₂.

C. Transfections

Environ 450 000 cellules QT6 ont étéensemencées sur des plaques 6 puits et transfectées 24 h après avec 2 µg de plasmide total à l'aide de JetPrime (Polyplus Transfection TM). Après 48 h, l'ADN total a été extrait à l'aide du kit EPICENTRE. Les cellules HeLa et 293T-UGI ont étéensemencées à 8 x 10⁵ cellules par puits sur des plaques 6 puits. Après 24h de culture, elles ont été transfectées avec 2 µg de plasmide total à l'aide de JetPrime (Polyplus Transfection TM). Les cellules ont été lysées après 48h de transfection. Pour la microscopie, 50 000 cellules HeLa ont étéensemencées sur des lames Lab-Tek et

Plasmid	Matrix	Primers
A3BDT196HI	A3Bi7	For : 5'-TACCTGATGGATCCAC CACAT ATTCACTTTCAACTTT-3' Rev : 5'-AAAGTTGAAAGTGAAT TATGTGT GGATCCATCAGGTA-3'
A3BF200S	A3Bi7	For : 5'-CCAGACACATTCACT TCCA ACTTTAATAATGAC-3' Rev : 5'-GTCATTATTAAAGTT TGGA GTGAATGTGTCTGG-3'
A3BΔPLV	A3Bi7	For : 5'-TTCAACTTTAATAATGACCTTCGACGGCGCCAGACC-3' Rev : 5'-GGTCTGGCGCCGTCGAAGGTCATTATTAAAGTTGAA-3'
A3BDPLVLR205GIG	A3Bi7	For : 5'-TTCAACTTTAATAAT GGCAT TGGAC GGCGCCAGACCTAC-3' Rev : 5'-GTAGGTCTGGCGCCG TCCA ATGCC ATTATTAAAGTTGAA-3'
A3BRQ212HK	A3Bi7 A3BDT196HI	For : 5'-TTGGTCCTTCGACGGC CATA AGACCTACTTGTGCTAT-3' Rev : 5'-ATAGCACAAAGTAGGT CTTATG CCGTCGAAGGACCAA-3'
A3BWVL228SVK	A3Bi7	For : 5'-CTGGACAATGGCACCT TCGGTCA AGATGGACCAGCACATG-3' Rev : 5'-CATGTGCTGGTCCAT CTTGACCG AGGTGCCATTGTCCAG-3'
A3BM235R	A3Bi7	For : 5'-CTGATGGACCAGCACAGGGGCTTTCTATGCAAC-3' Rev : 5'-GTTGCATAGAAAGCC CTGTG CTGGTCCATCAG-3'
A3BC239H	A3Bi7	For : 5'-CACATGGGCTTTCTACACAACGAGGCTAAGAAT-3' Rev : 5'-ATTCTTAGCCTCGTT TGTG TAGAAAGCCCATGTG-3'
A3BE241Q	A3Bi7	For : 5'-GGCTTTCTATGCAACC AGG CTAAGAATCTTCTCTGT-3' Rev : 5'-ACAGAGAAGATTCTT AGC CTGGTTGCATAGAAAGCC-3'
A3BCNE239HNQ	A3Bi7	For : 5'-CACATGGGCTTTCTACACAACC AGG CTAAGAATCTTCTCTGT-3' Rev : 5'-ACAGAGAAGATTCTT AGC CT GGTTG TGTAGAAAGCCCATGTG-3'
A3BEY342KH	A3Bi7	For : 5'-ACCTACGATGAGTTT AAGCA CTGCTGGGACACCTTT-3' Rev : 5'-AAAGGTGTCCAGCAG TGCTT AAACTCATCGTAGGT-3'
A3BYR350DH	A3Bi7	For : 5'-TGGGACACCTTTGTGG ACCACC AGGGATGTCCCTTC-3' Rev : 5'-GAAGGGACATCCCT GGTGGT CCACAAAGGTGTCCCA-3'
A3BE363D	A3Bi7	For : 5'-CCCTGGGATGGACTAGATGAGCACAGCCAAGCC-3' Rev : 5'-GGCTTGGCTGTGCTC ATCT AGTCCATCCCAGGG-3'
A3BcR212H	A3Bc	For : 5'-TTGGTCCTTCGACGGC CATC AGACCTACTTGTGCTAT-3' Rev : 5'-ATAGCACAAAGTAGGT CTGATG CCGTCGAAGGACCAA-3'
A3BcQ213K	A3Bc	For : 5'-TTGGTCCTTCGACGGC CGCA AGACCTACTTGTGCTAT-3' Rev : 5'-ATAGCACAAAGTAGGT CTTGC GCCGTCGAAGGACCAA-3'

Tableau 3 : Liste des amorces employées pour la mutagenèse dirigée d'A3B.

transfectées 24h après avec 1 µg de plasmide total à l'aide de l'agent de transfection FuGENE HD Transfection Reagent (Roche Applied Science).

D. Western blot

Les cellules ont été lavées au PBS et resuspendues dans un tampon de lyse (0.5% Nonidet P-40, 20 mM Tris-HCl, pH 7.4, 120 mM NaCl, et 1 mM EDTA) supplémenté avec un inhibiteur de protéase (Complete Protease Inhibitor, Roche Applied Science). Les lysats cellulaires ont été incubés 20 min sur glace et clarifiés par centrifugation à 14 000g pendant 30 min. Les protéines extraites ont été séparées par d'électrophorèse SDS-PAGE sur un gel 4-12% NuPAGE Bis-Tris dans un tampon de migration MOPS (Invitrogen). Les protéines ont ensuite été transférées sur membrane de nitrocellulose. Les protéines fusionnées V5 ont été mises en évidence par les techniques d'immunoblot standard. Les anticorps utilisés sont un anticorps monoclonal de souris anti-V5 conjugué à la HRP (Invitrogen) et un anticorps monoclonal de souris spécifique pour la β -actine conjugué à la HRP (Sigma).

E. Immunofluorescence

Après avoir lavé les cellules au PBS, celles-ci ont été fixées à l'aide de paraformaldéhyde 4% durant 20 min à température ambiante. Les cellules ont ensuite été perméabilisées avec un mélange 50/50 de méthanol/acétone durant 20 min à température ambiante. Les anticorps de souris anti-V5 (Invitrogen) ont été utilisés dilués au 1/200 durant 1h à température ambiante. Les cellules ont ensuite été incubées avec des anticorps couplés à l'Alexa 488 ou Alexa 555 (Invitrogen) dilués au 1/200 durant 1h à température ambiante et à l'obscurité. Les lames ont ensuite été montées avec le milieu de montage Vectashield-DAPI. L'imagerie a été réalisée sur un microscope Leica-TCS-SP5 équipé d'un objectif X1.2 WATER.

F. Mesure de désamination in vitro

Après 72h de transfection, les 293T transfectées ont été lavées au PBS et récoltées. Les protéines ont été extraites par sonication à l'aide d'un tampon de lyse spécifique (25 mM HEPES pH7.4, 10% glycérol, 150 mM NaCl, 0.5% Triton X-100, 1 mM EDTA, 1 mM MgCl₂, 1 mM ZnCl₂) supplémentés d'inhibiteurs de protéase. L'activité désaminase a été mesurée en incubant le lysat cellulaire avec 1pmole d'oligonucléotides couplés au FAM et au

TAMRA (5'-(6-FAM)-AAATTCTAATAGATAATGTGA-(TAMRA)-3') et en présence de 0,4 unités d'UDG (NEB) dans un tampon de réaction 20 mM Tris-HCl, 1 mM DTT, 1 mM EDTA. Après 2h d'incubation à 37°C, les sites abasiques générés ont été clivés à 95°C durant deux minutes et la fluorescence a ensuite été mesurée à l'aide d'un Realplex2 Mastercycler (BioRad). Les résultats ont été normalisés par rapport à la quantité totale de protéine à l'aide du kit Pierce BCA Protein Assay Kit (Thermo Scientific).

G. 3DPCR

Toutes les amplifications ont été décrites lors d'une étude précédente (Suspène, Aynaoud, Guétard, et al., 2011; Vartanian et al., 2010) avec 1 U de *Taq* polymérase (Eurobio) par réaction.

H. Analyse par cytométrie en flux des DSB

Après 48h de transfection, les cellules ont été lavées avec du PBS, fixées avec du paraformaldéhyde froid à 2-4% (Electron Microscopy Sciences) durant 10 min et perméabilisées avec 90% de méthanol froid (Sigma) durant 30 min. Après lavage au PBS, les cellules ont été incubées avec un anticorps anti-V5 de souris (Life Technologies) dilué au 1/200 dans un mélange PBS-BSA 0,5% pendant 1h. Après lavage au PBS, les cellules ont été incubées avec un anticorps IgG (H+L) Alexa Fluor 633 fragment F(ab')₂ anti-souris de chèvre (Life Technologies) durant 45 min. Les DSB ont été marqués à l'aide d'un anticorps monoclonal anti-γH2AX de lapin conjugué à l'Alexa Fluor 488 (Cell Signaling) dilué au 1/50. Toutes les incubations ont été faites sur glace. Les échantillons marqués ont été analysés sur une machine FACSCalibur en utilisant CellQuest Pro (BD Biosciences, version 5.2). Les données ont été analysées à l'aide du logiciel FlowJo (Tree Star Inc. version 8.7.1).

I. RNAseq sur les échantillons de sperme

Les spermatozoïdes de trois individus ont été sélectionnés et séparés des cellules somatiques par centrifugation de 1 à 2 ml de sperme à l'aide d'un gradient de Percoll à deux couches (45% et 90%) à 400g durant 20 min. La fraction de la couche à 90% a été lavée avec 2ml de Tyrode (Eurobio) et centrifugée à 400 g durant 10 min. Le culot a été récolté avec soin

afin d'éviter une contamination par les cellules somatiques puis a été lavé deux fois avec du PBS et l'ARN a été extrait à l'aide du kit RNeasy® mini kit (Qiagen). L'AND génomique a été éliminé par digestion sur colonne à la DNase avec le kit RNase-Free DNase Set (Qiagen). Les ARN ont ensuite été soumis à un second traitement DNase (TURBO™ DNase, Ambion Inc).

L'analyse RNAseq a été réalisée à l'aide du kit SOLiD™ Whole Transcriptome analysis kit for whole transcriptome libraries (Applied biosystems™) selon les recommandations du fabricant.

Les lectures de séquence gColor Space ont été cartographiées et quantifiées à l'aide de l'ABI Bioscope 1.3 Whole Transcriptome Analysis pipeline et du logiciel d'analyse Integromics SeqSolve et de scripts perl personnalisés. Les séquences de référence employées sont l'UCSC hg19 repeat-masked genome sequence et le NCBI RefSeq transcriptome dataset (September 2010). Un indice de pureté a été calculé à l'aide d'un script perl conçu par le laboratoire pour chaque échantillon. Celui-ci se définit par le ratio de chaque contaminant potentiel en fonction de l'intensité des transcrits spécifiques du sperme PRM1 et PRM2.

J. Analyse structurale

Les figures ont été réalisées à l'aide de Pymol (<http://www.pymol.org>). Les pKa ont été calculés à l'aide de PROPLA 3.0 sur les trois premiers modèles (dans l'ordre d'apparition) du fichier de coordonnées PDB 2M65. Le potentiel électrostatique a été calculé à l'aide d'APBS (Baker, Sept, Joseph, Holst, & McCammon, 2001) avec un contour d'une portée de -5 à 5kT/e. La mutagenèse a aussi été réalisée sur Pymol en utilisant les rotamers les plus communs et en évitant les collisions.

III- Résultat 3

A. Cellules et échantillons

Les PMBC obtenus de donneurs sains anonymes ont été isolés par gradient Ficoll (Eurobio). L'isolation des CD4⁺ a été réalisée à l'aide d'anticorps couplés à des billes magnétiques (Miltenyi Biotec). La pureté des CD4⁺ a été vérifiée par cytométrie en flux (FACSCalibur, Becton Dickinson) et est supérieure à 90%.

Les lymphocytes T CD4⁺ ont été stimulés avec 10µg/ml de PHA (Sigma, St-Louis, MO, USA), 100U/ml d'IL-2 (Sigma) et 500 U/ml d'IFNα (PBL Biomedical Laboratories, Piscataway, NJ, USA) durant 48h. Pour la détection d'hypermutations par 3DPCR, des lymphocytes T CD4⁺ ont été transduits avec le lentivirus rV2.EF1.UGI qui code une séquence optimisée de l'UGI, sous le contrôle du promoteur constitutif EF1 (Vectalys, Toulouse, France).

B. PCR et 3DPCR

Un clone moléculaire correspondant aux 376pb chevauchant l'exon 8 et l'intron 9 du gène *TP53* a été amplifié à partir de cellules HeLa. Environ 1ng d'ADN a été utilisé pour l'amplification de *TP53*. Les paramètres de la première étape de PCR sont : 5 min à 95°C puis 40 cycles (30 sec à 95°C, 30 sec à 60°C et 2 min à 72°C), et pour finir une étape de 10 min à 72°C. Les amorces utilisées sont : P53ext5 (5'-GAGCTGGACCTTAGGCTCCAGAAAGGACAA-3') et P53ext3 (5'-GCTGGTGTGTTGGGCAGTGCTAGGAA-3'). Les amplicons ont ensuite été dilués au 1/50 afin de réaliser la seconde étape de PCR sur une thermocycleur Eppendorf Mastercycler ep Gradient S (Eppendorf AG, Hamburg, Germany) programmé pour générer un gradient de dénaturation de 4°C. Le programme utilisé est le suivant : 5 min à 86-90°C puis 42 cycles (4 min à 86-90°C, 30 sec à 58°C et 2 min à 72°C), et enfin 10 min à 72°C. Le tampon de réaction employé se compose de 2.5mM MgCl₂, 16mM (NH₄)₂SO₄, 67mM Tris-HCl (pH 8.8), 0.01% Tween-20, 200µM de chaque désosynucleoside triphosphate, 100 mM de chaque amorce et 1,5U d'ADN polymérase *Taq* (EurobioTaq+; EurobioAbCys) ou de *Pfu* (PfuUltra II Fusion HS DNA Polymerase; Agilent Technologies, Santa Clara, CA, USA).

La β -caténine a été amplifiée selon le programme suivant : 5 min à 95°C puis 35 cycles (30 sec à 95°C, 30 sec à 60°C et 10 min à 72°C) et une étape finale de 20 min à 72°C. Les amorces employées sont : 5' β catout (5'-AGCTGATTTGATGGAGTTGGACA-3') et 3' β catout (5'-CCAGCTACTTGTTCTTGAGTGAA-3'). La PCR nichée a été réalisée avec 1/50 des amplicons de la première étape. Le programme utilisé est le suivant : 80-87°C durant 5 min puis 35 cycles (80-87°C durant 30 sec, 30 sec à 60°C et 10 min à 72°C) et une étape finale de 20 min à 72°C. Les amorces utilisées sont les suivantes: 5' β catin (5'-ACATGGCCATGGAACCAGACAGA-3') et 3' β catin (5'-GTTCTTGAGTGAAGGACTGAGAA-3').

Pour la PCR *long-range*, la première étape de réaction consiste en une dénaturation 5 min à 95°C suivie de 40 cycles (95°C durant 30 sec, 30 sec à 60°C et 10 min à 72°C), puis une élongation finale à 72°C de 20 min. Le tampon de réaction utilisé est composé de 2.5mM MgCl₂, 1X LA PCR Buffer II (TaKaRa LA Taq; TaKaRa, Otsu, Japan), 400 mM de chaque désoxynucleoside triphosphate (TaKaRa LA Taq; TaKaRa), 100 mM de chaque amorce et 1.5U de Taq (EurobioTaq+; Eurobio AbCys).

IV-Résultats 4

A. Réactifs

L'etoposide et le d'UTP proviennent de la société Sigma, l'actinomycine D de Millipore, l'inhibiteur de l'ARN polymérase III (ML-60218) de Merck Millipore, les dNTP de Fermentas, l'IFN α de PBL Biomedical Laboratories, le CpG d'Invitrogen, le polyIC et le Petidoglycan d'InvivoGen et enfin la Pfu d'Agilent Technologies. Les anticorps STING (antibody #3337), phospho-IRF-3 (Ser396) (4D4G, rabbit mAb #4947), IRF-3 5D6I4C, XP rabbit mAb #11904), β -catenin (antigène amino-terminal, #9581), anti IgG de lapin (produit chez la souris, L27A9, mAb #3678) et anti IgG de lapin couplé à la HRP (#7074) sont fournis par Cell Signaling Technology. L'anticorps RIG-I (E-5, sc-376882) est fourni par Santa Cruz Biotechnology, Inc. L'anticorps anti IgG de souris couplé à la HRP (#NA931V) est fourni par GE Healthcare. L'anticorps monoclonal anti β -actine couplé à la HRP (anticorps #A3854) provient de chez Sigma Aldrich.

B. Plasmides

Les ADNc codant A3A à A3H ont été amplifiés par PCR standard (*Taq* Platinum, Invitrogen) et clonés dans le vecteur pcDNA3.1/V5-His TOPO à l'aide du kit pcDNA3.1/V5-His TOPO Expression Kit (Invitrogen)

C. Culture cellulaire et transfection

Environ 450 000 cellules QT6 ont étéensemencées sur des plaques 6 puits et transfectées 24 h après avec 2 μ g de plasmide total à l'aide de JetPrime (Polyplus Transfection TM). Après 48 h, l'ADN total a été extrait à l'aide du kit EPICENTRE.

La lignée THP-1 (ATCC ® TIB-202TM) est maintenue en culture dans un milieu RPMI (Invitrogen) supplémenté avec 50U/ml de pénicilline, 50 μ g/ml de streptomycine, 10% de sérum de veau foetal à 37°C et 5% de CO₂. Pour la transfection, 1,5 x10⁶ cellules ont étéensemencées sur une plaque 12 puits. Après 24h de culture, les cellules ont été transfectées avec l'agent de transfection JetPRIME (Polyplus transfection). Après 24h de transfection, les surnageants ont été récoltés et l'IFN α (VeriKineTM Human IFN Alpha Multi-subtype ELISA Kit) et l'IFN β (VeriKineTM Human IFN Beta Multi-subtype ELISA Kit) ont été dosés et

analysés (PBL Assay Science). Tous les ADN ont été extraits de la même façon que précédemment.

Les cellules P2 humaines sont des cellules B transformées par EBV provenant de patients déficients pour l'UNG (Imai et al., 2003) et cultivé comme précédemment (Suspène, Aynaud, Guétard, et al., 2011). L'apoptose a été induite en incubant les cellules avec 100 μ M d'actinomycine D ou 100 μ M d'etoposide durant 16h. Les cellules ont ensuite été utilisées pour l'extraction d'ADN et le tri cellulaire

D. PBMC et isolation des lymphocytes T CD4+

Les PBMC obtenus de donneurs anonymes sains ont été isolés par gradient Ficoll (Eurobio). L'isolation des CD4+ a été réalisée à l'aide d'anticorps couplés à des billes magnétiques (Miltenyi Biotec). La pureté des CD4+ a été vérifiée par cytométrie en flux (FACSCalibur, Becton Dickinson) et est supérieure à 90%. Deux millions de lymphocytes T CD4+ ont été ensemencés sur plaques 24 puits et cultivées 24h avant tri cellulaire.

E. Tri cellulaire

Les cellules P2 ont été récoltées 16h après traitement à l'actinomycine et l'etoposide. En parallèle, les cellules P2 non traitées ont été récoltées. Toutes les cellules ont été resuspendues dans 500 μ l de PBS et les cellules viables ont été triées sur une plaque PCR 96 puits (ABgene, Thermo Scientific) à l'aide du MoFlo (Beckman Coulter). Les plaques ont ensuite été immédiatement refermées à l'aide d'un film adhésif PCR (Thermo Scientific) et stockées à -20°C.

F. Relargage du cytochrome c

Seize heures après le traitement par des drogues pro-apoptotiques, le relargage du cytochrome c par les cellules P2 a été étudié à l'aide du kit FlowCelect Cytochrome c Kit de Millipore selon les recommandations du fournisseur. Les échantillons marqués ont été analysés sur une machine FACSCalibur en utilisant CellQuest Pro (BD Biosciences, version 5.2). Les données ont été analysées à l'aide du logiciel FlowJo (Tree Star Inc. version 8.7.1).

G. Analyse de l'apoptose par cytométrie.

Après 16h de traitement avec les drogues pro-apoptotique, les cellules P2 ont été resuspendues dans un tampon de "binding" (BD Pharmigen) et marquées à l'aide d'un anticorps Annexin V couplé au FTIC (1 µg/ml) (BD Pharmigen). Afin de différencier les cellules en apoptose précoce des cellules en apoptose tardive ou en nécrose, les cellules ont été marquées avec 5µg/ml d'iodure de propidium (BD Pharmigen). L'analyse a été réalisée comme décrit précédemment.

H. Analyse des cassures de l'ADN double brin par cytométrie en flux.

Après 24h de transfection, les THP-1 ont été lavées avec du PBS, fixées avec du paraformaldéhyde froid à 2% (Electron Microscopy Sciences) durant 10 min et perméabilisées avec 90% de méthanol froid (Sigma) durant 30 min. Après lavage au PBS, les cellules ont été incubées avec un anticorps monoclonal anti-γH2AX de lapin conjugué à l'Alexa Fluor 488 (Cell Signaling) dilué au 1/50. Toutes les incubations ont été faites sur glace. Les échantillons marqués ont été analysés sur une machine MACSQuant Analyzer (Miltenyi Biotec) et le logiciel MACsQuantify™ (Miltenyi Biotec) ou FlowJo software (Tree Star, Inc., version 8.7.1). Pour chaque échantillon, 10 000 évènements ont été comptés.

I. PCR/3DPCR

Les échantillons provenant de patients ont été décrits dans une étude précédente (ref). La première étape d'amplification pour le gène *MT-COI* humain ou de caille comprend une étape de 5 min à 95°C suivi de 40 cycles (95°C ou 85°C durant 30 secondes, 60°C durant 30 secondes et 2 min à 72°C) et finalement une étape de 10 min à 72°C. La seconde étape de PCR et la 3DPCR ont été réalisées avec 1µl du produit de réaction de la première PCR. Les paramètres pour la PCR standard sont: 95°C durant 5 min puis 40 cycles (30s à 95°C, 30s à 55 ou 60°C et 2 min à 72°C), complété par une étape de 10 min à 72°C. La seconde étape de 3DPCR a été réalisée comme décrit précédemment, mais avec un gradient de dénaturation de 81 à 87°C. Pour l'amplification du gène humain *MT-CYB*, la première étape de PCR comprend une étape de 5 min à 95°C, suivie de 40 cycles (30 sec à 95°C, 30 sec à 60°C et 2 min à 72°C), et enfin une étape de 10 min à 72°C. Pour la seconde étape de 3DPCR, les paramètres sont les mêmes que pour le gène *MT-COI*. Les réactions de PCR ont été réalisées avec 2,5 U de *Taq* polymérase (Bioline). Les amorces utilisées ainsi que l'amplification des

régions V1V2 du VIH-1 suivent le protocole décrit lors d'une étude précédente (Suspène, Aynaud, Guétard, et al., 2011; Suspène, Aynaud, Vartanian, & Wain-Hobson, 2013; Suspène, Henry, et al., 2005).

J. PCR quantitative

Les ARN totaux des cellules P2, THP-1 et des lymphocytes T CD4⁺ ont été extraits à l'aide du kit RNeasy Plus Mini Kit (Qiagen) selon les recommandations du fabricant. Les ADNc ont ensuite été synthétisés à l'aide du kit Quantitec Reverse Transcription Kit à partir d'un µg d'ARN et selon le protocole fourni par le fabricant.

Les niveaux d'expression des *A3* et de *RPL13A* ont été quantifiés par qPCR à l'aide d'un Eppendorf Realplex2. Les réactions ont été réalisées sur une plaque 96 puits contenant pour chacun des puits 12.5 µL de 2x probe Master Mix (Roche), 2 µl d'amorces (5 µm chacune), 0.33 µL d'UPL probe (Roche), 5.17 µL d'eau et 5 µl d'ADN). Le programme employé est le suivant: 95°C 10min suivi de 45 cycles (95°C 15 sec, 58°C 15 sec, 68°C 30 sec). Les amorces employées pour la quantification du niveau d'expression des *A3* et de *RPL13A* sont celles décrites par (Refsland et al., 2010). Les résultats sont normalisés par rapport au niveau d'expression de *RPL13A*.

K. Western Blot

Les protéines ont été extraites après 24h de transfection. Les Western Blot ont été réalisés selon les protocoles standards en employant des anticorps monoclonaux de lapin ou de souris. Après incubation avec un anticorps anti IgG de souris couplé à la HRP (Amersham), la membrane a été révélée par chemiluminescence (Pierce).

L. Co-immunoprecipitation ARN/protéine.

Après 8 heures de transfection avec 500ng d'ADN V1V2 de VIH-1, 1.5×10^6 , les THP-1 ont été incubés avec 37% de formaldéhyde (concentration finale de 1%) durant 15 min à température ambiante sous agitation douce. Le "cross-linking" a été interrompu par l'ajout de glycine 2M pour une concentration finale de 0,2M durant 5 min à température ambiante. Les cellules ont été centrifugées à 1800 g durant 5 min à 4°C et lysées sur la glace dans un tampon de lyse (#9803, Cell Signaling) contenant 40U/ml de RNasin. Les lysats ont été soniqués 3 fois sur la glace, 5 secondes chaque fois. Après une centrifugation de 10 min à 14,000 g à

4°C, les surnageants ont été transférés dans des nouveaux tubes et incubés 1 heure à 4°C avec 30 µl de billes couplés à la protéine A (#9863, Cell Signaling). Après 10 min de centrifugation à 14 000 g à 4°C, les surnageants ont été transférés et incubés avec un anticorps anti-RIG-I en rotation orbital à 4°C sur la nuit. Trente microlitres de billes d'agarose/Protéine A ont été ajoutés et incubés sous rotation à 4°C pendant 30 min. Après centrifugation à 14 000 g à 4°C pendant 10 min, les culots ont été lavés cinq fois avec 500µL de tampon de lyse 1X contenant 40U/ml de RNasin. Les culots ont ensuite été resuspendus dans 50µL d'eau et chauffés à 95°C durant 2 min. Après une ultime centrifugation de 10 min à 4°C et à 14 000 g, une RT-PCR a été réalisée sur le surnageant.

RESULTATS

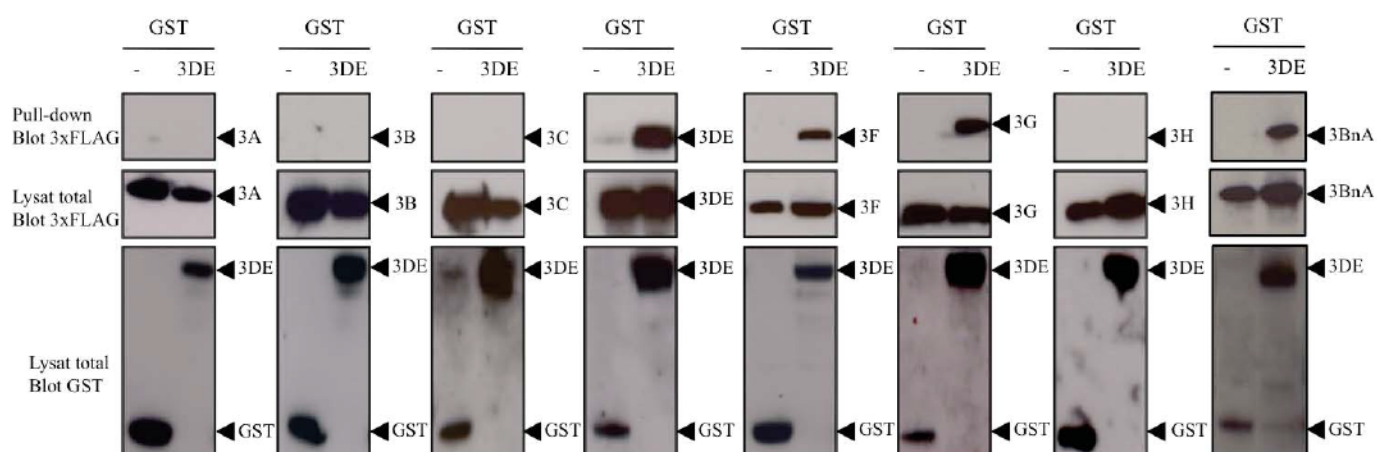


Figure 24 : A3DE dimérise avec elle même, A3F et A3G.

Des cellules HEK293T ont été co-transfectées avec une construction GST-A3DE ou la GST seule comme contrôle négatif et A3A, A3B, A3C, A3DE, A3F, A3G, A3H ou la chimère A3Bna avec un tag 3xFlag en N-terminal. Les GST pull-down ont ensuite été suivis de Western blots avec un anticorps anti-3xFlag.

I. Résultat 1 : Régulation de l'*editing* du Virus de l'Hépatite B par hétérodimérisation des cytidines désaminases APOBEC3

A. Introduction

Le locus *APOBEC3*, situé sur le chromosome 22, est composé de 7 gènes qui codent une famille de sept cytidines désaminases (A3A-A3C, A3DE, A3F-A3H) impliquées dans les défenses antivirales (Jarmuz et al., 2002). Celles-ci désaminent les cytidines situées sur l'ADN simple brin, préférentiellement dans le contexte 5'TpC (Bishop et al., 2004; Suspène, Aynaud, Guétard, et al., 2011), à l'exception d'A3G qui désamine dans le contexte 5'CpC (Beale et al., 2004; Suspène et al., 2004). Les A3 inhibent la réplication de nombreux virus tels que le VIH (Harris et al., 2003; Lecossier, Bouchonnet, Clavel, & Hance, 2003; Mangeat et al., 2003; Mariani et al., 2003; Sheehy et al., 2002), les herpèsvirus (Suspène, Aynaud, Koch, et al., 2011), les papillomavirus (Vartanian et al., 2008; Z. Wang et al., 2014; Warren et al., 2015) mais aussi de nombreux rétroéléments (Bogerd et al., 2006; H. Chen et al., 2006; Muckenfuss et al., 2006). Leur action se traduit par la présence de nombreuses mutations de type CG->TA, appelées *editing*, sur les génomes des virus impactés (Harris et al., 2003; Lecossier et al., 2003; Mangeat et al., 2003; Mariani et al., 2003; Suspène et al., 2004). A3A et A3G font par ailleurs partie des ISG (*Interferon stimulating genes*) (Aynaud et al., 2012; Bonvin et al., 2006; Koning et al., 2009; Peng, Lei, Jin, Greenwell-Wild, & Wahl, 2006; Stenglein et al., 2010). A3A, A3C, A3F, A3G et A3H peuvent induire des mutations sur l'ADN mitochondrial cytoplasmique, tandis qu'A3A et A3B peuvent induire des mutations sur l'ADN chromosomique (Burns, Lackey, et al., 2013; Caval, Suspène, Shapira, et al., 2014; Mussil et al., 2013; Shinohara et al., 2012; Suspène, Aynaud, Guétard, et al., 2011). Ces deux enzymes désaminent les bases 5-méthylcytidines (Carpenter et al., 2012; Caval, Suspène, Shapira, et al., 2014; Suspène et al., 2013; Wijesinghe & Bhagwat, 2012) mais seule A3A peut induire des DSB (Caval, Suspène, Vartanian, et al., 2014; Landry et al., 2011; Mussil et al., 2013).

Ces enzymes peuvent être composées d'un domaine catalytique unique (A3A, A3C et A3H) ou de deux domaines catalytiques (A3B, A3DE, A3F et A3G) mais dont seul le domaine C-terminal est actif (Hache, Liddament, & Harris, 2005; Jarmuz et al., 2002;

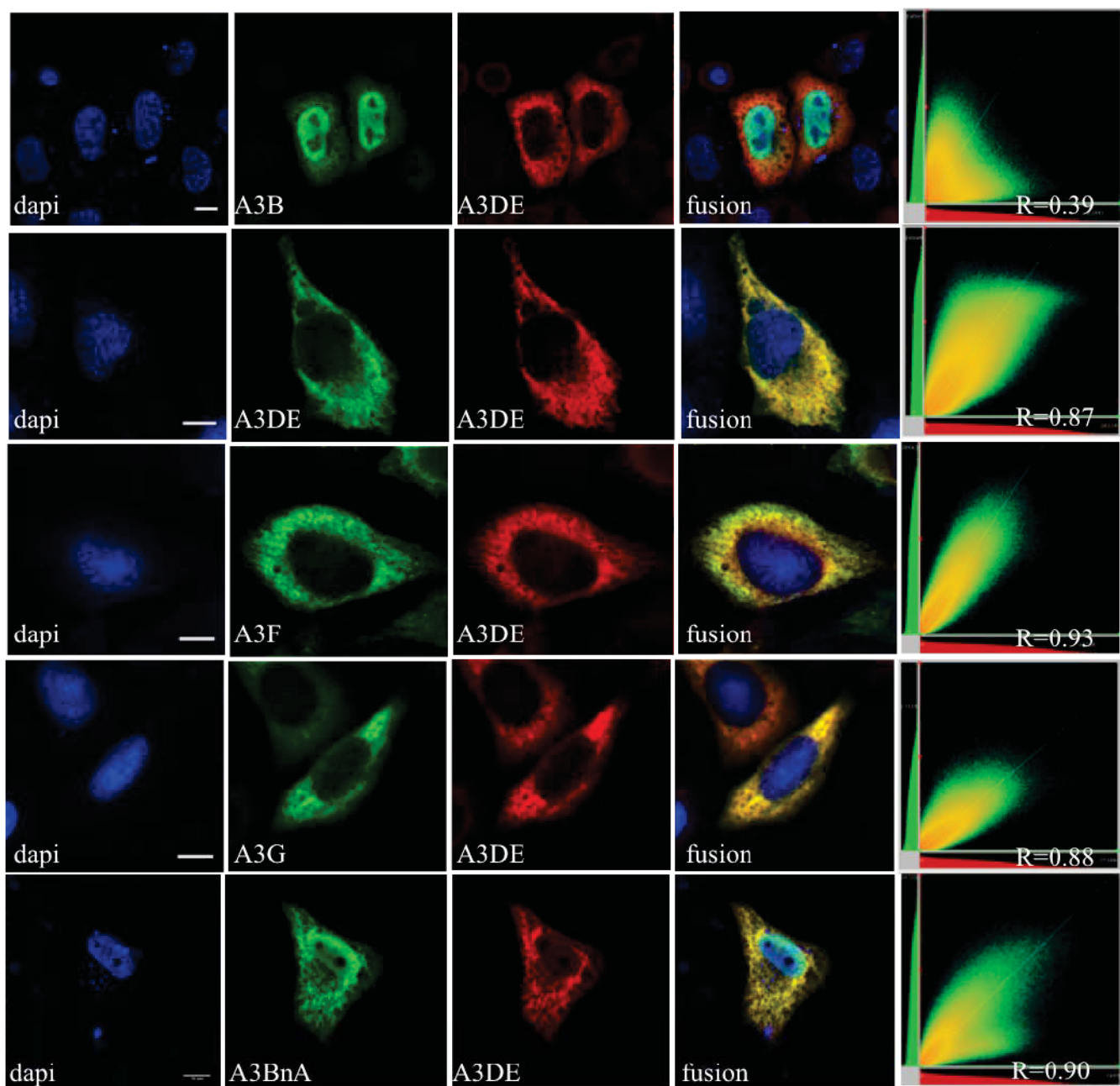


Figure 25 : Microscopie confocale des A3DE avec d'autres protéines A3.

Après 36h de transfection dans les HeLa, A3DE-V5, A3F-V5 et A3BnA semblent, contrairement à A3B-V5, fortement co-localiser avec 3xFlag-A3DE. Le diagramme et le coefficient de corrélation (R) renseignent sur la significativité de l'interaction.

Newman et al., 2005). Ces domaines sont composés d'une région en doigt de zinc permettant la fixation à l'ADN(Jarmuz et al., 2002). Par ailleurs, le domaine N-terminal inactif dans le cas des A3 à double domaine améliore l'efficacité de l'enzyme à s'associer à l'ADN(Henry et al., 2009). Il a été montré que de nombreuses A3 peuvent former des homo- et des hétérodimères (Bohn et al., 2015; Friew, Boyko, Hu, & Pathak, 2009; Huthoff, Autore, Gallois-Montbrun, Fraternali, & Malim, 2009; Koyama et al., 2013; Stauch et al., 2009; Wiegand et al., 2004), bien qu'AID, A3A et A3G soient actives *in vitro* en monomères (Aynaud et al., 2012; Brar et al., 2008; Salter, Krucinska, Raina, Smith, & Wedekind, 2009). Enfin, A3G, qui est l'un des principaux facteurs de restriction contre le VIH-1 et le VHB, peut aussi former des complexes de haut poids moléculaire (Aynaud et al., 2012; Brar et al., 2008; Y. L. Chiu et al., 2006; McDougall et al., 2011; Salter et al., 2009; Wedekind et al., 2006).

A3DE a la particularité de ne pas posséder d'activité catalytique du fait de la fixation d'une mutation Tyr320Cys dans le domaine C-terminal. La mutation vers Tyr320 permet de restaurer totalement l'activité cytidine désaminase contre le VIH-1(Dang et al., 2011). Néanmoins, A3DE participe de façon limitée à la restriction des rétroéléments ou de virus indépendamment de l'*editing* (Holmes, Koning, Bishop, & Malim, 2007; Newman et al., 2005; Nguyen et al., 2007; Rosler et al., 2005). A3DE, à l'instar d'A3F et d'A3G est strictement cytoplasmique et est composée de deux domaines. D'un point de vu phylogénétique, le domaine N-terminal d'A3DE (A3D) est proche du domaine N-terminal d'A3F (A3Fn), et dans une moindre mesure d'A3B (A3Bn) et d'A3G (A3Gn). Son domaine C-terminal (A3E) est proche de celui d'A3F (A3Fc) et dans une moindre mesure d'A3C (Conticello, Thomas, Petersen-Mahrt, & Neuberger, 2005; Henry et al., 2009; LaRue, Jonsson, et al., 2008).

A3G constitue le principal facteur de restriction du VHB, éditant jusqu'à 35% des génomes *in vivo* (Vartanian et al., 2010). A3DE, A3F et A3G sont co-exprimés dans de nombreux tissus et surexprimés dans les cirrhoses (Dang et al., 2006; Koning et al., 2009; Peng et al., 2007). Ceci est intrigant étant donné qu'A3DE possède peu voire aucune activité cytidine désaminase (Chaipan, Smith, Hu, & Pathak, 2013; Dang et al., 2006; Duggal, Malik, & Emerman, 2011). Sachant qu'A3B, A3F et A3G peuvent former des homo- et hétérodimères, nous avons étudié le rôle d'A3DE dans le contexte de l'infection au VHB en nous intéressant à ses interactions avec les autres A3. Nous montrons dans ce travail

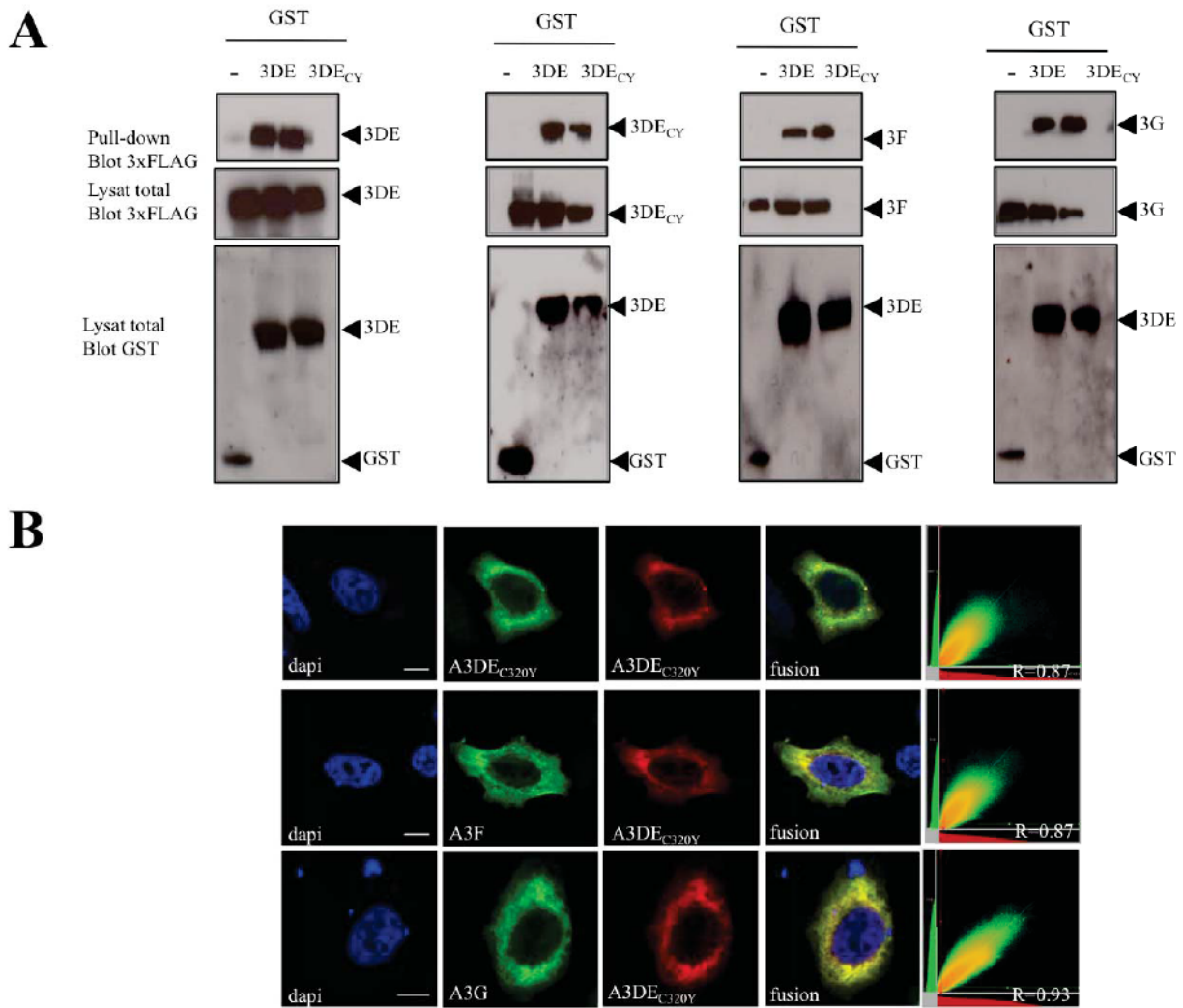


Figure 26 :A3DE_{C320Y} dimérise et colocalise avec lui même, A3DE, A3F et A3G.

A) Les cellules HEK293T ont été co-transfectées avec les constructions GST-A3DE, GST-A3DE_{C320Y} ou un plasmide GST vide et avec les constructions 3xFlag-A3DE, 3xFlag-A3DE_{C320Y}, A3F et A3G. Un GST-Pull-down suivi d'un Western Blot ont ensuite été réalisés.

B) Co-localisation d'A3DE_{C320Y}-V5, A3F-V5 et A3G-V5 avec 3xFlag-A3DE_{C320Y} dans les cellules HeLa 36 h après transfection. Le diagramme et le coefficient de corrélation (R) renseignent sur la significativité de l'interaction.

qu'A3DE inhibe l'activité d'A3F le VHB et dans une moindre mesure A3G en formant des hétérodimères avec celles-ci.

B. Résultats

1) APOBEC3DE forme des dimères avec APOBEC3F et APOBEC3G

Afin d'étudier les interactions qu'A3DE peut avoir avec les autres A3, nous avons réalisé une expérience de GST pull-down. Cette technique permet d'isoler des complexes protéiques en purifiant les protéines fusionnées à la GST (A3DE ici) et en mettant en évidence par Western Blot les interacteurs portant un tag différent (3xFlag ici). Comme nous pouvons l'observer sur la figure 24, A3DE interagit fortement avec elle-même, avec A3F et A3G mais ne forme pas de complexe avec A3A, A3B, A3C et A3H. L'absence d'interaction avec A3B est surprenante étant donné que les domaines A3D et d'A3Bn partagent 79% d'identité au niveau protéique, ce qui est bien plus élevé qu'entre A3D et A3Gn (43%) qui eux interagissent. Afin d'en déterminer la raison, nous avons réalisé un GST pull-down avec A3DE et une chimère A3B/A fonctionnelle (A3BnA) codant le domaine N-terminal d'A3B fusionné avec la séquence d'A3A à la place du domaine C-terminal d'A3B, qui diffère seulement de 9% au niveau protéique. On observe qu'A3DE interagit avec l'hybride fonctionnel A3BnA (Figure 24).

Dans le but de confirmer les résultats obtenus par GST pull-down, nous avons analysé la localisation d'A3DE et de ses partenaires par microscopie confocale (Figure 25). Une co-localisation forte et significative est observée pour A3DE avec elle-même, A3F et A3G mais pas avec A3B (Figure 25), ni avec A3A, A3C et A3H (résultats non présentés). Néanmoins, on observe qu'A3DE co-localise avec A3BnA (Figure 25).

Les résultats négatifs pour A3B et A3A mais positifs pour A3BnA montrent que l'interaction avec A3DE implique le domaine N-terminal d'A3B tandis que la séquence C-terminal d'A3B semble empêcher cette interaction. De nombreux résidus ont été décrits comme étant déterminants pour la dimérisation d'A3G (Bennett, Salter, Liu, Wedekind, & Smith, 2008; Chelico, Prochnow, Erie, Chen, & Goodman, 2010; Shlyakhtenko et al., 2011; Wedekind et al., 2006). Nous avons réalisé des mutations analogues sur le domaine N-terminal d'A3DE.

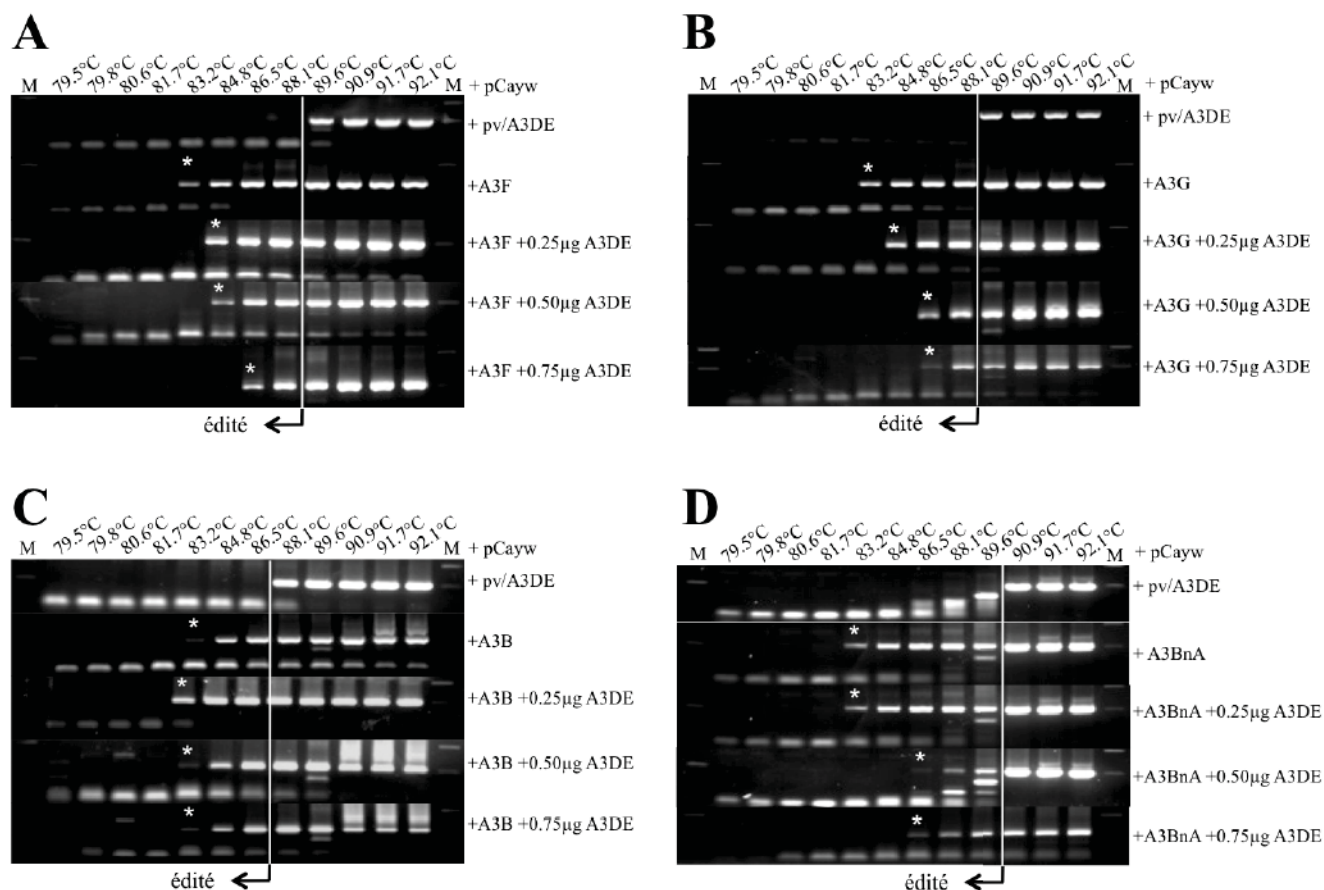


Figure 27 : A3DE protège l'ADN du VHB contre la désamination par les A3 de manière dose-dépendante.

Les cellules QT6 ont été co-transfectées avec le plasmide pCayw, A3F (A), A3G (B), A3B (C) ou A3BnA (D) et des quantités croissantes d'A3DE (0,25 μg ; 0,5 μg ; 0,75 μg) et complétés avec la quantité nécessaire de plasmide vide pour arriver à 2 μg total. Les lignes délimitent les produits de 3DPCR édités. Les astérisques indiquent les amplicons clonés et séquencés. M : marqueur de poids moléculaire.

Ces mutations, Tyr28Ala, Trp94Ala, Tyr124Ala et Tyr126Ala, ne suffisent pas à abolir l'interaction d'A3DE avec A3F ou A3G (données non présentées).

Etant donné que la présence d'une cystéine en position 320 perturbe la fixation d'A3DE à l'ADN simple brin, nous avons testé la capacité du mutant actif d'A3DE, A3DE_{C320Y}, à interagir avec A3F et A3G. Nous pouvons observer par GST-pulldown qu'A3DE_{C320Y} interagit avec elle-même, A3DE_{C320}, A3F et A3G (Figure 26A). L'analyse par microscopie confocale indique qu'A3DE_{C320Y} se comporte comme A3DE_{C320} puisqu'il colocalise avec lui-même, A3DE_{C320}, A3F et A3G (Figure 26B). Ces résultats indiquent qu'il n'y a pas de lien entre la perte d'activité d'A3DE et sa capacité à former des hétérodimères ou avec l'absence d'interaction avec A3B.

2) APOBEC3DE inhibe l'*editing* du VHB par APOBEC3F et APOBEC3G

Pour étudier l'impact de l'interaction d'A3DE avec A3F et A3G sur la réplication de l'ADN du VHB, nous avons co-transfectés la lignée QT6 avec le plasmide pCayw (un clone moléculaire du VHB), A3F ou A3G et des quantités croissantes d'A3DE. La lignée QT6, dérivée de fibroblastes de cailles, est utilisée car elle n'exprime pas d'A3 endogène et permet de s'affranchir du bruit de fond que l'on aurait eu en employant une lignée cellulaire humaine (Henry et al., 2009; Suspène, Aynaud, Guétard, et al., 2011; Vartanian et al., 2010). Après 48h de transfection, l'ADN total a été extrait et amplifié par 3DPCR, technique permettant d'amplifier spécifiquement l'ADN édité par les A3 en utilisant un gradient de températures de dénaturation (Suspène, Henry, et al., 2005). Comme nous pouvons l'observer sur la figure 27A et 27B, la présence de concentration croissante d'A3DE entraîne une modification de la Td minimale d'amplification de façon dose dépendante, la dernière amplification étant indiquée par un astérisque. En accord avec les expériences de GST pull-down et de colocalisation, A3DE n'a pas d'impact sur l'*editing* du VHB par A3B (Figure 27C). Néanmoins, A3DE inhibe l'*editing* du VHB par A3BnA (Figure 27D). Les échantillons indiqués par un astérisque ont été clonés et séquencés et présentent tous la signature de l'*editing* par les A3, avec une préférence pour le contexte 5'TpC pour A3B, A3BnA et A3F et le contexte 5'CpC pour A3G (données non présentées).

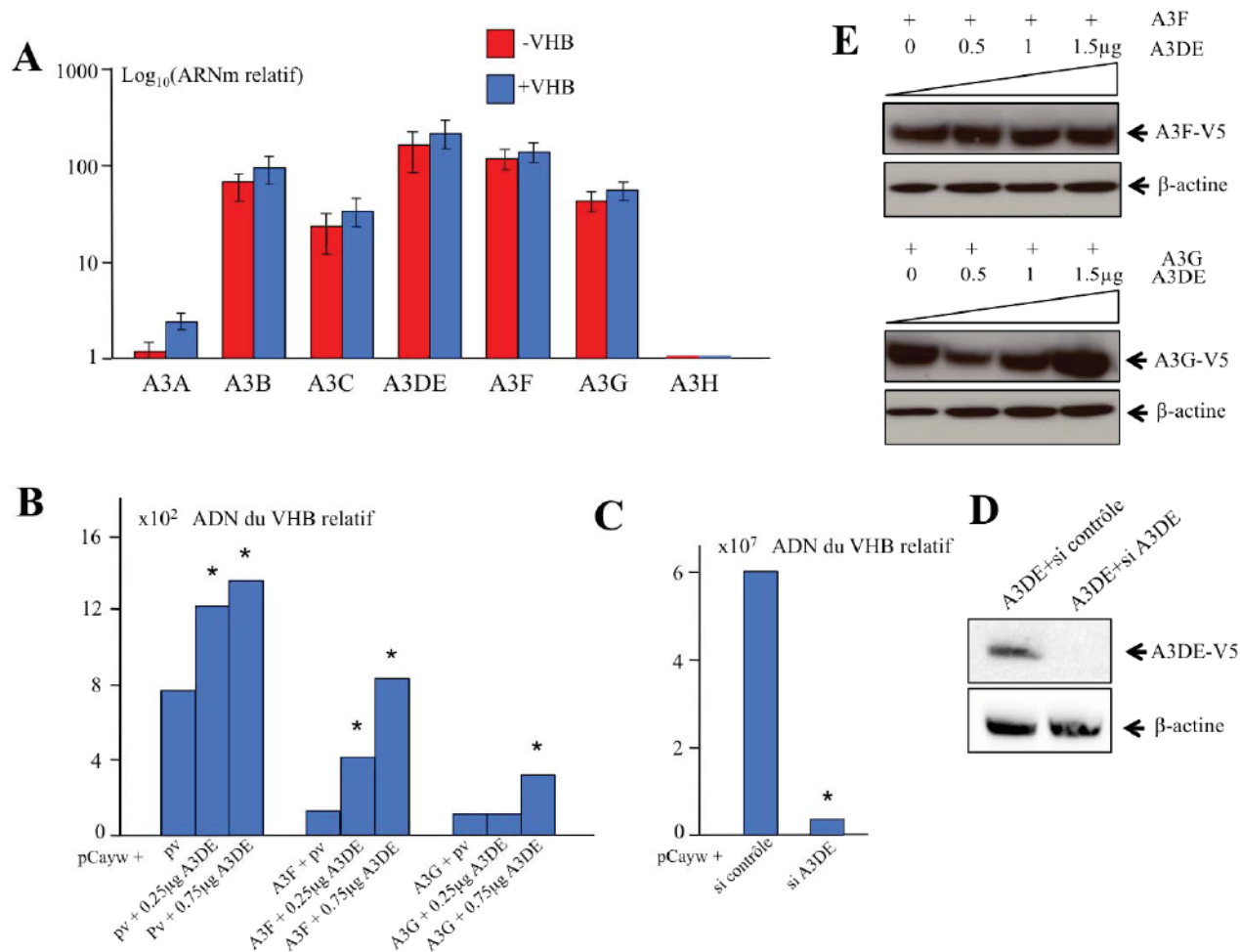


Figure 28 : L'interaction d'A3DE avec A3F et A3G entraîne leur inhibition et l'augmentation des niveaux d'ADN de VHB.

A) Profil transcriptionnel des *APOBEC3* dans la lignée Huh7 transfectée ou non transfectée avec le VHB. Les quantités ont été normalisées par rapport au niveau du gène de référence *RPL13A*. B) La lignée Huh7 a été transfectée avec le plasmide pCayw ou le plasmide vide ou A3F ou A3G et 0,25 μ g ou 0,75 μ g d'A3DE. A 7 jours après transfection, l'ADN total du surnageant a été extrait et une PCR quantitative sur le gène *X* du VHB a été réalisée. C) Les cellules Huh7 ont été transfectées avec le plasmide pCayw et un siARN contrôle ou dirigé contre A3DE. Après 48h de transfection, l'ADN total des surnageants a été extrait et une PCR quantitative similaire à celle réalisée en B) a été faite. D) Test d'efficacité du siARN A3DE 48h après transfection des siARN et d'A3DE-V5 par Western Blot. E) Les cellules Huh7 ont été transfectées avec A3F ou A3G et des quantités croissantes d'A3DE (quantité totale d'ADN de 2 μ g). Le niveau *steady-state* d'A3F et d'A3G a été déterminé par Western blot à 24h. La β -actine sert ici de témoins de charge.

3) APOBEC3DE promeut la réplication du VHB

Afin de mesurer l'expression relative des A3 dans la lignée hépatique Huh7, nous avons réalisé une PCR quantitative par Taqman sur ces cellules en présence ou en absence de VHB (Figure 28A). En comparant l'expression des A3 au gène de référence *RPL13A*, nous observons que les cellules Huh7 expriment bien plus d'A3B, d'A3C, d'A3DE, d'A3F et d'A3G que d'A3A et d'A3H. De façon intéressante, l'expression des A3 ne varie pas en fonction de la présence ou l'absence de VHB. Ceci suggère que les protéines virales n'induisent pas d'expression des A3. De plus, nous observons une proportion similaire d'expression d'A3DE, d'A3F et d'A3G, montrant que le niveau endogène d'A3DE n'est pas suffisant pour totalement inhiber A3G et A3F.

Afin de quantifier l'impact d'A3DE sur l'expression du VHB, nous avons co-transfecté la lignée Huh7 avec le plasmide pCayw et avec des quantités croissantes d'A3DE et quantifié par PCR quantitative l'ADN du VHB dans le surnageant. Comme nous pouvons l'observer sur la Figure 28B, 0,25µg et 0,75µg d'A3DE entraînent l'augmentation de la quantité d'ADN du VHB dans le surnageant de culture, vraisemblablement en inhibant les A3 endogènes. Lorsque A3F ou A3G sont co-transfectés avec le VHB, les niveaux d'ADN de VHB dans le surnageant diminuent d'un facteur 5 à 7 (Figure 28B). La transfection de quantités croissantes d'A3DE dans ces conditions permet d'inhiber la restriction par A3F, et dans une moindre mesure par A3G (Figure 28B), en accord avec l'homologie de séquences entre les domaines N-terminaux d'A3DE et A3F (67%) et ceux d'A3DE et d'A3G (44%). A l'inverse, l'emploi d'un siARN A3DE dans la lignée Huh7 entraîne une restriction du VHB 20 fois plus importante que lors de l'emploi d'un siARN contrôle (Figure 28C). L'efficacité du siARN A3DE pour le *silencing* d'A3DE est présentée dans la figure 28D. Malgré l'inhibition d'A3F et d'A3G par A3DE, la formation d'hétérodimères n'affecte pas la stabilité des protéines A3F et A3G (figure 28E). Ceci indique que l'inhibition par A3DE ne se fait pas par dégradation d'A3F et d'A3G mais probablement par encombrement du site actif, empêchant ainsi leur action.

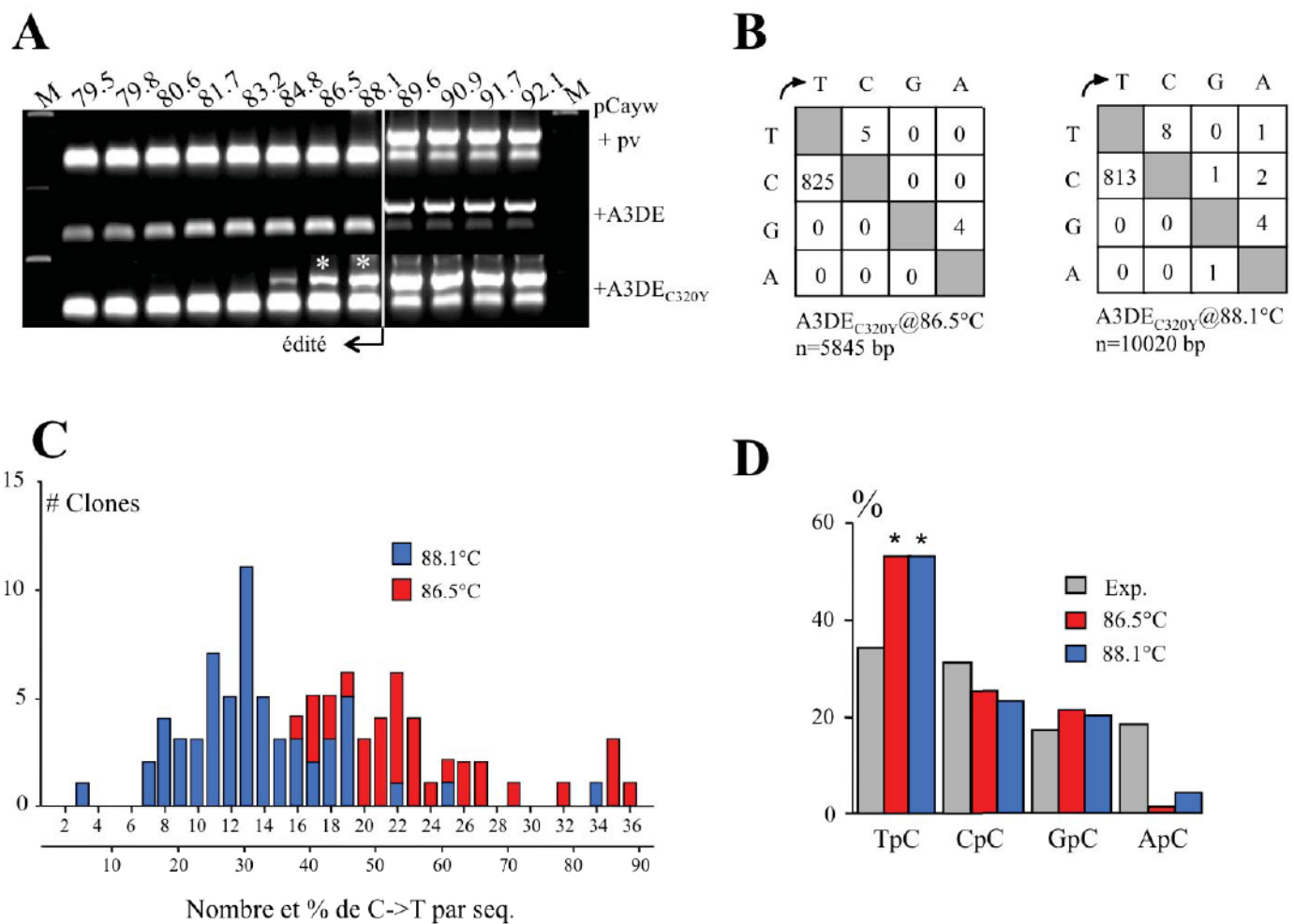


Figure 29 : A3DE_{C320} n'est pas fonctionnel mais peut acquérir une activité par la mutation C320Y

A) 3DPCR sur le gène *X* du VHB. La lignée de caille QT6 a été transfectée avec le plasmide pCayw et A3DE, A3DE_{C320Y} ou un plasmide vide en tant que contrôle négatif. Après 72 heures de transfection, l'ADN total a été extrait et une 3DPCR a été réalisée sur le gène *X* du VHB. La taille du fragment 3DPCR est de 203 pb. Les astérisques indiquent que le produit PCR a été cloné et séquencé. M : marqueur de poids moléculaire. B) Matrices de mutations de séquences de VHB hyperéditées à deux températures différentes. Le chiffre sous les matrices indique le nombre de bases séquencées. C) Distribution des fréquences d'*editing* par clone à 86,5°C et 88,1°C. # indique le nombre de clones séquencés. D) Analyse du contexte dinucléotidique 5' des résidus cytidines édités. Exp : valeurs attendues. Un test χ^2 a été réalisé et les différentes significatives sont indiquées par un astérisque (*, $p < 0,005$).

4) APOBEC3DEC320Y hyperédite le génome du VHB.

Il a auparavant été suggéré qu'A3DEC₃₂₀ inhibe la réplication du VIH-1 et éventuellement induise des mutations de l'ADN simple brin (Dang et al., 2006; Hultquist et al., 2011; K. Sato et al., 2014). Néanmoins, aucune séquence hyperéditée n'appuie ces hypothèses et seules des fréquences d'*editing* d'environ 0,02% sont observées (Dang et al., 2006), plus de 1000 fois moins que pour des hypermutants classiques. Afin d'étudier ceci, nous avons de nouveau employé la technique de 3DPCR, suffisamment sensible pour la détection d'hypermutants puisqu'elle permet l'amplification d'une molécule d'ADN éditée parmi 10⁴ molécules non éditées (Suspène, Henry, et al., 2005). La transfection du VHB et d'A3DEC₃₂₀ dans la lignée QT6 ne permet pas de mettre en évidence d'*hyperediting* sur le gène *X* du VHB. En revanche, lorsque A3DEC_{320Y} est transfectée, des produits de 3DPCR pour le gène *X* sont obtenus jusqu'à 84,8°C, contre 89,6°C pour A3DE et le contrôle plasmide vide (figure 29A). Les produits de 3DPCR des températures 88,1°C et 86,5°C ont été clonés. Des mutations de VHB édité de façon extensive et monotone sont obtenues, avec des taux de cytidines mutées allant jusqu'à 90% (Figure 29B, D). Le contexte dinucléotidique obtenu est uniforme avec un fort biais pour le contexte 5'TpC, typique d'A3A, A3B, A3C et A3H (Figure 29C, (Beale et al., 2004; Bishop et al., 2004; Suspène, Aynaud, Guétard, et al., 2011)), suivi par le contexte 5'GpC qui est inhabituel pour ces enzymes. De façon intéressante, l'*editing* par 3F de la même région suit le même biais (TpC>GpC>CpC>ApC), ce qui est en accord avec ce que l'on observe physiologiquement (Henry et al., 2009). De plus, A3DEC_{320Y} induit aussi des mutations sur l'ADN mitochondrial cytoplasmique avec le même biais dinucléotidique, indiquant qu'A3DEC_{320Y} n'a pas besoin d'être encapsidée pour éditer l'ADN simple brin (données non présentées).

Ces résultats montrent qu'A3DEC₃₂₀ n'induit pas d'hypermutations sur l'ADN du VHB (Figure 29C). Intrigués par cette observation et du rôle *a priori* pro-viral d'A3DE par inhibition de la restriction du VHB par A3F et A3G, nous avons génotypé 41 cas de cirrhoses du foie issus d'une étude précédente (Vartanian et al., 2010) et 60 échantillons provenant de carcinomes hépatocellulaires (CHC) associés au VHB, en plus de différentes lignées cellulaires dont des lignées de CHC (Tableau 1). Nous avons retrouvé la forme A3DEC₃₂₀ dans tous les échantillons analysés. La substitution C320Y correspond à un SNP G>A

Patients	
3 Sains	Foie
10 VHB ⁺	Cirrhose
10 VHB ⁺	Cirrhose
11 VHB ⁺ VHC ⁺	Cirrhose
10 Alcoolique	Cirrhose
10 VHB ⁺ Chine	CHC
10 VHB ⁺ Itale	CHC
10 VHB ⁺ Maroc	CHC
10 VHB ⁺ Perou	CHC
10 VHB ⁺ Roumanie	CHC
10 VHB ⁺ Tunisie	CHC
Cell lines	
Huh7	CHC
OC4CH16	CHC
HLF	CHC
NUK-1	CHC
BEL9204	CHC
IHH	Hépatocytes immortalisés
Fa2N	Hépatocytes immortalisés
HHEmi	Hépatocytes primaires
LSEC	Cellules endothéliales sinusoidales de foie
HT29F	Adénocarcinome du côlon
HCT116	Carcinome colorectal
HaCaT	Kératinocytes immortalisés
HEK293	Cellules de rein d'embryon

Tableau 1 : Analyse de l'allèle présent d'APOBEC3DE chez des patients et différentes lignées cellulaires

(rs369944754). L'étude de la base de donnée « *1000 genomes* » indique que l'allèle A3DE_{C320Y} n'est présent qu'à une fréquence très faible de ~0,1% (www.1000genomes.org) dénotant qu'A3DE n'est pas polymorphique sur cette position.

C. Conclusion

Nous avons, lors de cette étude, mis en évidence pour la première fois la fonction d'A3DE. Il est paradoxal que ce membre d'un cluster de gènes antiviraux participe à l'inhibition des membres de ce cluster, améliorant ainsi la réplication du VHB. De plus, A3DE est surexprimée dans les cirrhoses alcooliques ou viro-induites (Vartanian et al., 2010).

Bien que non totalement titrée, l'amélioration de la réplication du VHB par A3DE dans la figure 28A est d'environ 70%. Le polymorphisme A3DE_{C320Y} étant complètement fixé dans la population humaine, il semblerait que posséder l'allèle inactif confère néanmoins un avantage évolutif. Autrement, un important *bottleneck* évolutif pourrait être de cet allèle (Gagneux et al., 1999). Cet événement est récent dans l'évolution étant donné que seuls les gorilles en plus des Hommes portent une substitution similaire (C320). Toutes les autres espèces de singe portent l'allèle actif Y320, dont le macaque pour lequel il a été montré que l'A3DE est active (Duggal et al., 2011). La présence de la forme C320 chez l'Homme et le gorille conforte l'hypothèse que cet allèle confère un avantage dans certaines circonstances.

Avec la découverte de ces interactions entre A3DE, A3F et A3G, il apparaît que le rôle de certaines de ces enzymes de restriction est plus subtil que l'on ne l'imaginait jusqu'ici. Le promoteur proximal d'A3DE est le plus proche de celui d'A3C (53% d'identité sur les 700 premières bases). Ces deux promoteurs sont insensibles à l'interféron, contrairement à A3A et A3G qui sont des ISG (Aynaud et al., 2012; Bonvin et al., 2006; Koning et al., 2009; Peng et al., 2006; Stenglein et al., 2010). Il est possible que l'induction d'A3G par l'interféron permette de submerger l'expression basale d'A3DE et ainsi d'outrepasser l'inhibition. Etant donné qu'A3DE, comme les autres A3, est fortement exprimée dans la lignée hématopoïétique, il est possible qu'elle participe à abaisser le niveau seuil d'A3F et d'A3G, facilitant ainsi la fonction de Vif.

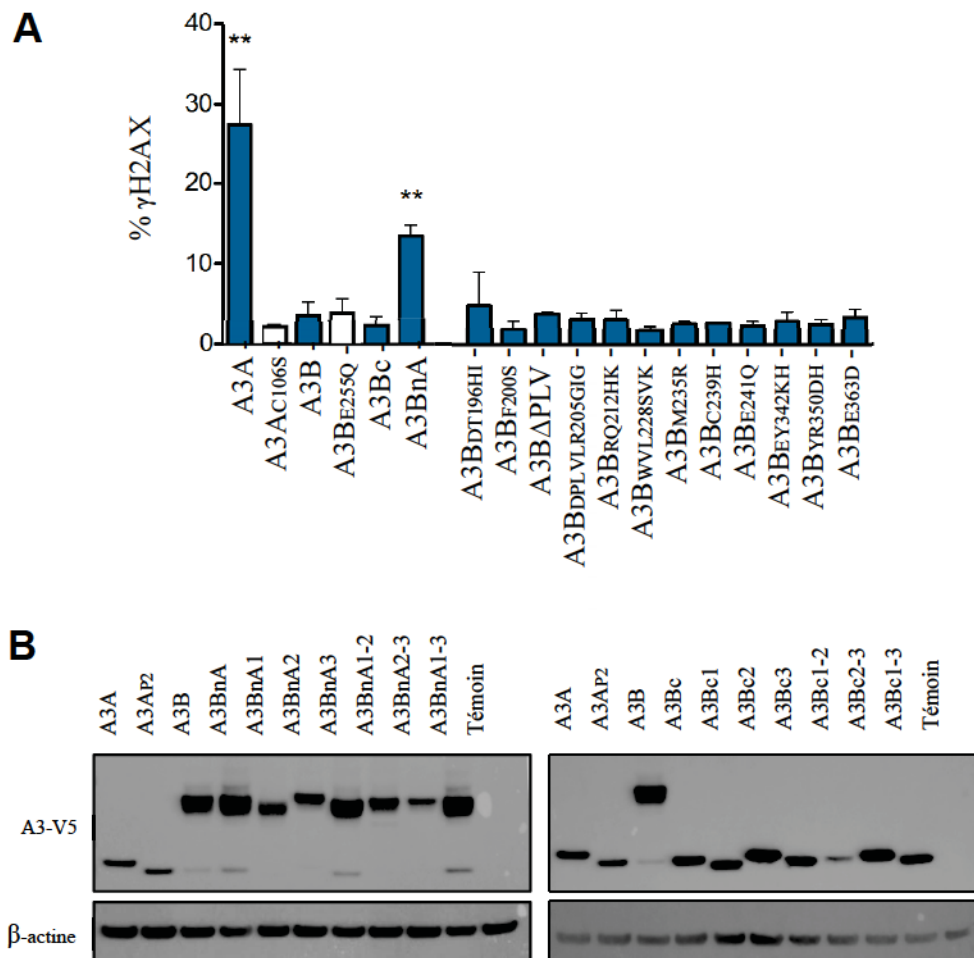


Figure 30 : Analyse par FACS des cellules positives pour γ H2AX et analyse par western blot de l'expression des chimères A3 taggées V5.

A) Analyse par FACS des cellules HeLa γ H2AX positives parmi les cellules V-5 positives 48h après transfection des mutants A3B. Les barres d'erreur sont calculées à partir de six transfusions indépendantes. Les tests de significativité sont calculés par rapport aux mesures réalisées avec le mutant catalytique A3AC106S en utilisant le test de Mann-Whitney (* $P < 0,1$, ** $P < 0,01$, *** $P < 0,001$). B) Analyse par Western blot de l'expression des chimères d'A3 humaines taggées V5 transfectées dans la lignée HEK293T. La β -actine sert ici de témoins de charge.

II. Résultat 2 : Bases moléculaires du phénotype atténué de l'enzyme mutatrice de l'ADN APOBEC3B

A. Introduction

Les gènes *APOBEC3A* et *APOBEC3B* codent deux cytidine désaminases, A3A et A3B, actives sur l'ADN simple brin (Caval, Suspene, Shapira, et al., 2014; Kostrzak, Henry, Demoyen, Wain-Hobson, & Vartanian, 2014; Stenglein et al., 2010; Suspène, Aynaud, Guétard, et al., 2011). Celles-ci sont d'importantes sources de mutations dans de nombreux génomes issus de cancers, dans lesquels on retrouve majoritairement des mutations de type CG->TA (Alexandrov, Nik-Zainal, Wedge, Aparicio, et al., 2013; Burns, Lackey, et al., 2013; Caval, Suspene, Shapira, et al., 2014; Greenman et al., 2007; Komatsu, Nagasaki, Fujimori, Amano, & Miki, 2008; Nik-Zainal et al., 2012; Suspène, Aynaud, Guétard, et al., 2011). Bien que ces deux enzymes induisent des hypermutations sur l'ADN chromosomique dans un contexte expérimental (Alexandrov, Nik-Zainal, Wedge, Aparicio, et al., 2013; Burns, Lackey, et al., 2013; Caval, Suspene, Shapira, et al., 2014; Greenman et al., 2007; Komatsu et al., 2008; Nik-Zainal et al., 2012; Suspène, Aynaud, Guétard, et al., 2011), seule A3A peut induire des DSBs malgré une divergence de seulement 9% entre A3A et A3B (Caval, Suspene, Shapira, et al., 2014). Nous nous sommes par conséquent proposé de déterminer les bases moléculaires à l'origine de l'atténuation de la génotoxicité d'A3B.

B. Résultats

1. Etude de l'impact des deux domaines d'A3B.

Afin de conférer à A3B la capacité d'induire des DSBs, nous avons généré par mutagenèse dirigée 12 mutants de la partie C-terminal d'A3B, sur la base des homologies trouvées avec A3A (Figure 30). Nous avons ensuite, après transfection, mesuré le taux de DSBs par FACS après marquage des histones γ H2AX. Aucune de ces mutations n'a suffi à conférer à A3B la capacité d'induire des DSBs (Figure 30). Ceci suggère que plusieurs

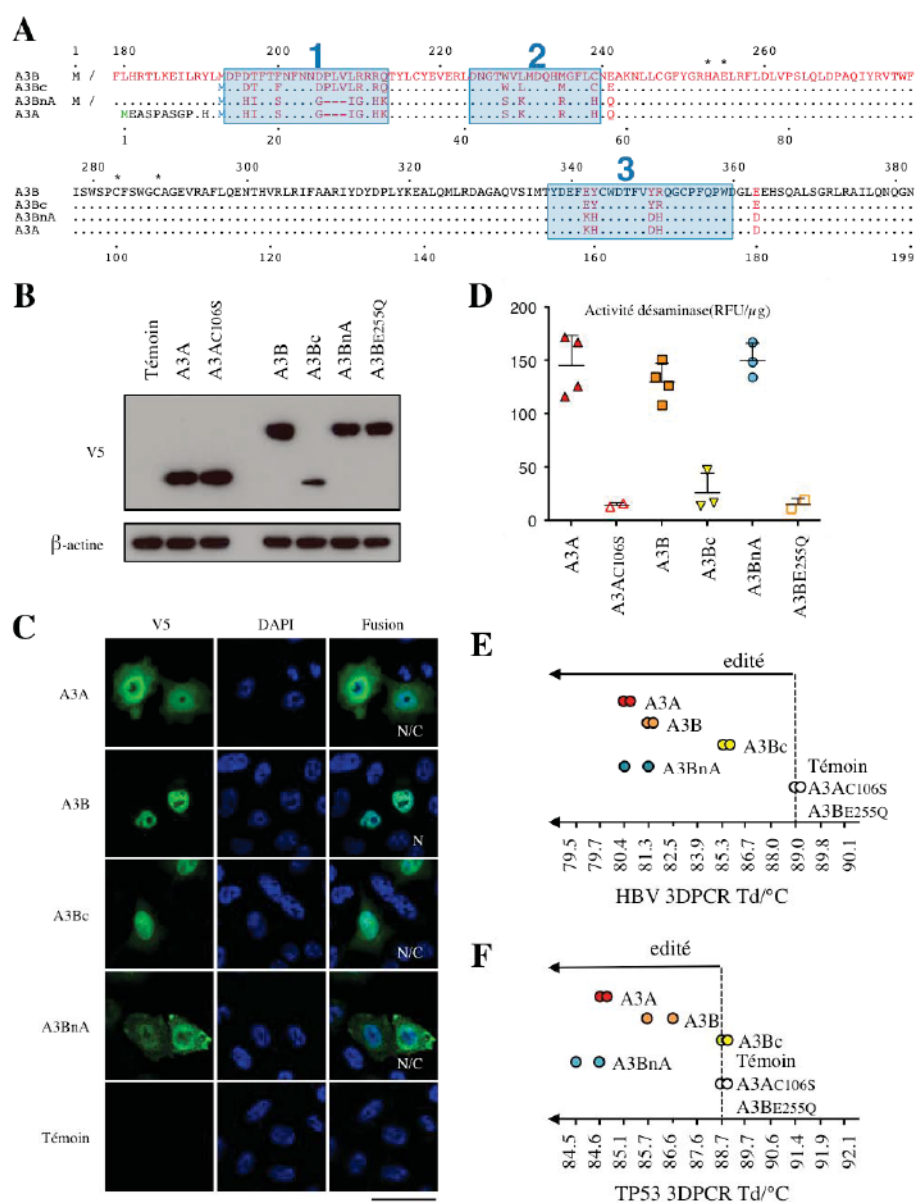


Figure 31 : Chimères A3A/A3B

A) Alignement CLUSTALW des séquences en amino-acide d'A3A, A3B, A3BnA et A3Bc. Les régions de mutation sont encadrées en bleu tandis que les résidus impliqués dans le domaine de coordination de zinc sont marqués par un astérisque. Les résidus variables sont colorés en rouge. La numérotation du haut correspond à celle d'A3A tandis que la numérotation du bas correspond à celle d'A3B. B) Western Blot des APOBEC3 taggées V5 dans la lignée HEK293T. Le β -actine sert ici de témoin de charge.

C) Localisation cellulaire des A3. Les A3 taguées V-5 sont observées au microscope confocal dans la lignée HeLa 24h post transfection. Les noyaux sont marqués au DAPI. N : Localisation nucléaire, N/C : localisation nucléo-cytoplasmique. D) Mesure de désamination *in vitro* sur un oligonucléotide couplé TAMRA-FAM de lysats issus de cellules 293T transfectées avec les A3. Le bruit de fond a été obtenu par la mesure de la désamination sur des lysats de cellules transfectées par un plasmide témoin. RFU : Relative Fluorescence Unit. E) 3D-PCR sur le génome du VHB après transfection de la lignée QT6 avec les protéines A3 F) 3D-PCR sur TP53 après transfection de la lignée 293T-UGI avec les protéines A3.

résidus sont impliqués dans l'atténuation d'A3B. Nous avons par conséquent envisagé une approche différente, consistant à générer plusieurs chimères A3A/A3B : A3BnA, composée du domaine N terminal d'A3B et d'A3A en domaine C-terminal et A3Bc constitué seulement du domaine C-terminal d'A3B. Cette dernière a été générée par l'emploi de l'équivalent du second codon Met retrouvé dans la séquence d'A3A (Figure 31A) (Mussil et al., 2013; Thielen et al., 2010). On observe par Western blot qu'A3BnA s'exprime correctement tandis qu'A3Bc s'exprime moins bien qu'A3A (Figure 31B). L'étude de la localisation de ces chimères par microscopie confocale montre que si A3Bc à l'instar d'A3A à une localisation nucléo cytoplasmique, A3BnA est bien plus localisée dans le noyau que dans le cytoplasme. Ces observations suggèrent que les deux domaines d'A3B contribuent à sa localisation strictement nucléaire (Figure 31C).

Afin d'évaluer l'activité cytidine désaminase de ces chimères, nous avons transfecté dans la lignée 293T les plasmides codant ces deux chimères couplées à un tag V5. Les cellules ont été lysées et une mesure de l'activité désaminase *in vitro* a été effectuée. Le principe de cette mesure est basé sur le FRET : une cible couplée au TAM et au FAM et possédant un site de désamination 5'TpC est mise en présence du lysat dans un tampon contenant de l'UNG : s'il y a désamination de la cible, l'UNG va exciser le U et ainsi entraîner un clivage de la cible et la libération du FAM fluorescent de son quencher, le TAM (Caval, Suspene, Shapira, et al., 2014; Stenglein et al., 2010).

Ainsi, on observe qu'A3BnA a une activité cytidine désaminase égale à A3A et A3B tandis qu'A3Bc semble très peu active (Figure 31D). Afin de confirmer ces résultats, nous avons co-transfecté ces chimères avec un plasmide codant le génome complet du VHB dans la lignée QT6. Nous observons par 3DPCR que les deux chimères sont actives. Néanmoins, A3Bn semble moins active qu'A3BnA, A3A et A3B (Figure 31E).

Nous nous sommes aussi intéressés à l'activité de ces chimères sur l'ADN nucléaire en les transfectant dans la lignée 293T-UGI (Caval, Suspene, Shapira, et al., 2014; Suspène, Aynaud, Guétard, et al., 2011). Nous avons ensuite employé la technique de 3DPCR sur l'ADN extrait de ces cellules. Nous observons qu'A3BnA est aussi active qu'A3A et qu'A3B, contrairement à A3Bc qui n'est pas du tout active (Figure 31F). Etant donné que la 3DPCR n'est pas une technique quantitative, nous avons mesuré la génotoxicité de ces protéines par quantification des DSBs par FACS. Nous observons qu'A3Bn et A3B n'induisent pas de DSBs contrairement à A3BnA qui induit des DSBs à un niveau modéré (50% moins que les

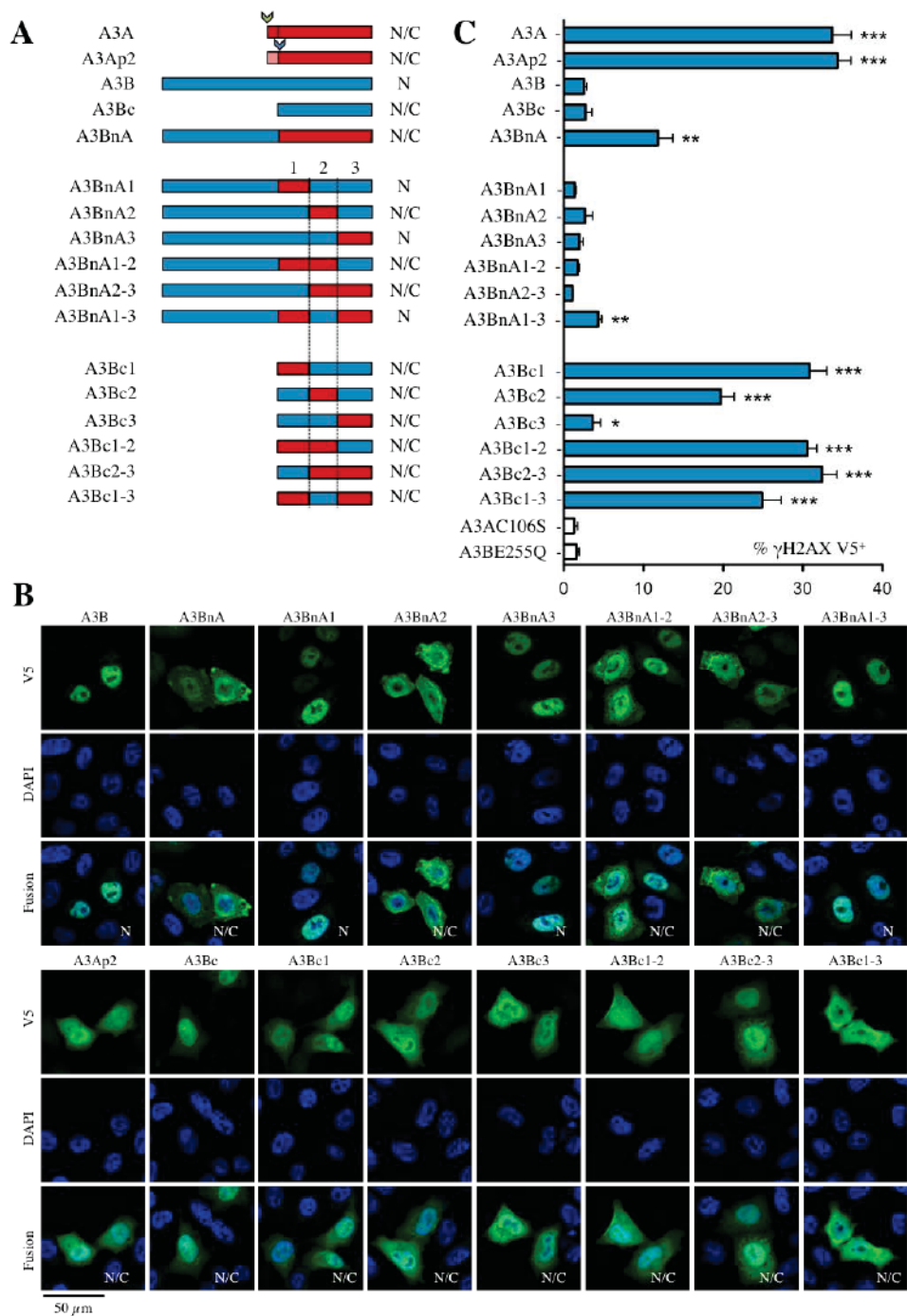


Figure 32 : Représentation schématique d'A3BnA et A3Bc et localisation cellulaire des A3

A) Représentation schématique des chimères dérivant d'A3BnA et d'A3Bc. B) Localisation cellulaire des A3. Les A3 taguées V-5 sont observées au microscope confocal dans la lignée HeLa 24h post transfection. Les noyaux sont marqués au DAPI. N : Localisation nucléaire, N/C : localisation nucléo-cytoplasmique. C) Analyse par FACS des cellules HeLa γ H2AX positives parmi les cellules V-5 positives 48h après transfection. Les barres d'erreur sont calculées à partir de six transfections indépendantes. Les tests de significativité sont calculés par rapport aux mesures réalisées avec le mutant catalytique A3A_{C106S} en utilisant le test de Mann-Whitney (*P<0,1, **P<0,01, ***P<0,001).

témoins positifs A3Ap1 et A3Ap2). Nous en concluons donc qu'A3Bn peut atténuer l'activité d'A3A et par conséquent être utilisée pour les expériences de pertes de fonction, tandis que A3Bc peut être utilisée pour le gain de fonction (Figure 32A, 32C).

Etant donné que les 18 résidus (9% de la séquence) distinguant A3Ap2 d'A3Bc sont localisés dans trois régions différentes (1, 2 et 3) (Figure 31A), nous avons généré six chimères A3BnA et A3Bc en substituant ces 3 régions par leur région équivalente d'A3A (Figure 32A). Le niveau d'expression de ces 12 protéines est similaire mais on observe des différences de localisation : A3BnA1, A3BnA3 et A3BnA1-3 ont une localisation exclusivement nucléaire suggérant que des résidus d'A3B présents dans la région 2 sont nécessaires pour la localisation nucléaire (Figure 32B et Figure 31B). Tandis que les mutations sur A3BnA conduisent quasi-systématiquement à une perte des DSBs, la mutation de la région 1 d'A3Bc (A3Bc1) permet en revanche de restaurer les DSBs à un niveau comparable à A3A et A3Ap2 (Figure 32C). La mutation de la seconde région permet aussi de restaurer les DSBs mais à un niveau moindre tandis que A3Bc3 à un effet faible mais significatif (Figure 32C). Les mutations des régions 2 et 3 combinées ont un effet comparable à A3Bc1 et aux contrôles positifs A3Ap1 et A3Ap2.

2. Les déterminants moléculaires de l'atténuation d'A3Bc

Afin de déterminer précisément les résidus permettant de restaurer l'activité DSBs chez A3Bc1 et A3Bc2, des combinaisons plus petites de résidus d'A3Ap2 ont été introduites dans la séquence d'A3Bc et l'activité DSBs a été mesurée (Figure 33A). De toutes les constructions réalisées, seul les substitutions RQ212HK et dans une moindre mesure DT196HI permettent de restaurer l'induction de DSBs de manière significative (Figure 33A). La combinaison de ces deux substitutions permet une meilleure activité DSBs mais demeure inférieure de moitié à l'activité d'A3Ap1 et A3Ap2 (Figure 33B). Ceci suggère que d'autres mutations de la région 1, bien que n'entraînant pas un gain de fonction seule, contribuent à la génotoxicité lorsqu'elles sont combinées aux mutations DT196HI et RQ212HK. Des mutations ponctuelles montrent que les substitutions Q213K, D196H et T197I sont les plus importantes, puisqu'elles permettent à elles seules de restaurer l'induction de DSBs juste au-dessus du bruit de fond (Figure 33B). De façon intéressante, la mutagenèse réciproque d'A3Ap2 montre que la mutation HK29RQ (numérotation d'A3A, Figure 33A) entraîne une

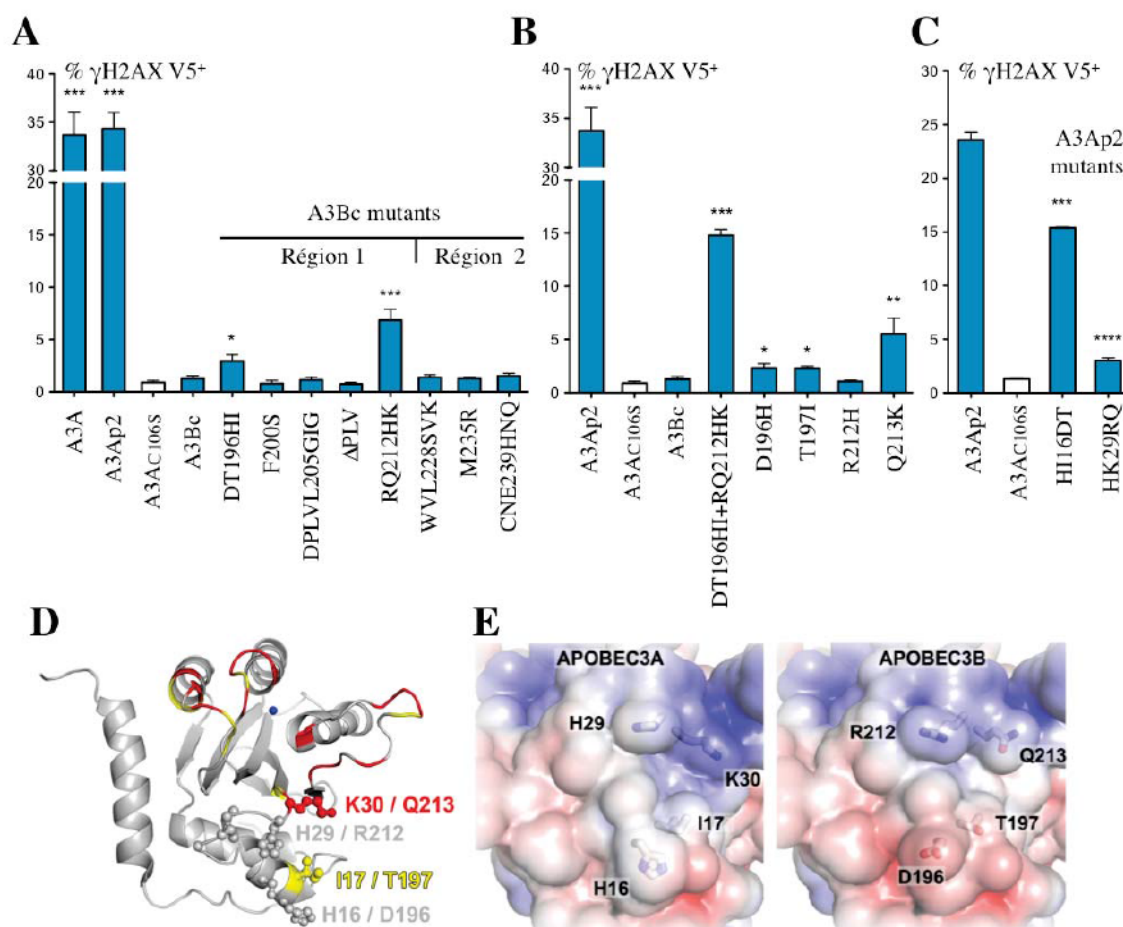


Figure 33 : Analyse par FACS des cellules HeLa γ H2AX positives et représentation de la structure d'A3A.

A) Analyse par FACS des cellules HeLa γ H2AX positives parmi les cellules V-5 positives 48h après transfection des mutants A3B. Les barres d'erreur sont calculées à partir de six transfactions indépendantes. Les tests de significativité sont calculés par rapport aux mesures réalisées avec le mutant catalytique A3AC106S en utilisant le test de Mann-Whitney (* $P < 0,1$, ** $P < 0,01$, *** $P < 0,001$). B) Analyse par FACS des mutants A3Bc simple et quadruple résidus. La figure est annotée de la même façon qu'en A. C) Analyse par FACS des mutants A3Ap2. La figure est annotée de la même façon qu'en A. D) Représentation « Cartoon » de la structure d'A3A. Les chaînes latérales des résidus identifiés par mutagenèse sont représentées en « ball-and-sticks ». Les résidus décrits comme subissant une modification chimique lors de la fixation d'un nucléotide sont représentés en rouge ($>0,05$ p.p.m) et jaune (0,028-0,0050 p.p.m). L'atome de zinc est représenté par une sphère bleue. La numérotation des acides aminés est indiquée selon la nomenclature d'A3A et d'A3B respectivement. E) Représentation en surface solide d'A3A et d'A3B. Le potentiel électrostatique est coloré en fonction de sa valeur, calculée à partir d'APBS (-5 kT/e [rouge] to 5 kT/e [bleu]). La différence en amino-acide entre A3A et A3B modifie le potentiel électrostatique de surface.

atténuation des DSBs plus importante que la mutation H116DT, faisant miroir avec l'effet des mutations sur A3Bc (Figure 33C).

L'analyse par NMR du binding d'A3A à l'ADN simple brin révèle que le résidu K30, et dans une moindre mesure, le résidu I17 (respectivement Q213 et T197 sur A3B, Figure 31A) subissent un changement chimique : ceci souligne leur participation à la fixation au substrat. Ces résidus participent à la fixation du squelette de l'ADN simple brin (Figure 33D, 33E).

3. L'atténuation d'A3Bc est conservée dans l'évolution

Contrairement à A3A qui est orthologue chez les mammifères placentaires, A3B ne serait apparue dans l'évolution qu'à partir des singes cercopithèques. L'analyse phylogénétique du couple A3A/A3Bc chez les primates montre que rhA3A et rhA3Bc du macaque rhésus forment un groupe séparé d'A3A ou A3Bc tels qu'elles sont présentes chez l'Homme, le chimpanzé et le bonobo, témoignant d'une conversion génétique récente (Figure 35). Parmi les 22 résidus (11% de la séquence) distinguant rhA3A et rhA3Bc, seulement six sont aussi différents entre A3A et A3Bc chez l'Homme (Figure 34A). Afin de voir si la même différence de génotoxicité existe chez le macaque rhésus, nous avons conçu quatre constructions rhA3A et rhA3B. Ceux-ci sont aussi bien exprimés et ont la même localisation que leur équivalent humain (Figure 34B, 5C). De la même façon que les enzymes A3 humaines, rhA3B et rhA3Bc, contrairement à rhA3A et rhA3AP2, n'induisent pas de DSBs (Figure 34D). La comparaison entre les résidus d'A3A et d'A3B de macaque correspondant aux mutations H116DT et K213Q identifiées plus tôt révèle seulement une mutation correspondant à un polymorphisme D/N déjà décrit dans la littérature (Figure 34A, encadrés) (Caval, Suspene, Vartanian, et al., 2014). Plus globalement, ces mutations, que ce soit chez l'Homme ou le macaque, contribuent à une acidification du domaine A3Bc comparé à A3A, ce qui mène à une affinité amoindrie pour l'ADN simple brin, celui-ci étant chargé négativement. De façon intéressante, cette observation peut être étendue à toutes les séquences d'A3Bc disponibles, domaine le plus acide de toutes les A3 (Figure 35).

Etant donné le caractère atténué mais conservé du phénotype d'A3B, nous avons voulu déterminer si cette enzyme peut contribuer à la variabilité génétique chez l'Homme. Pour cela, nous avons analysé les transcriptomes RNAseq issus d'échantillons de sperme

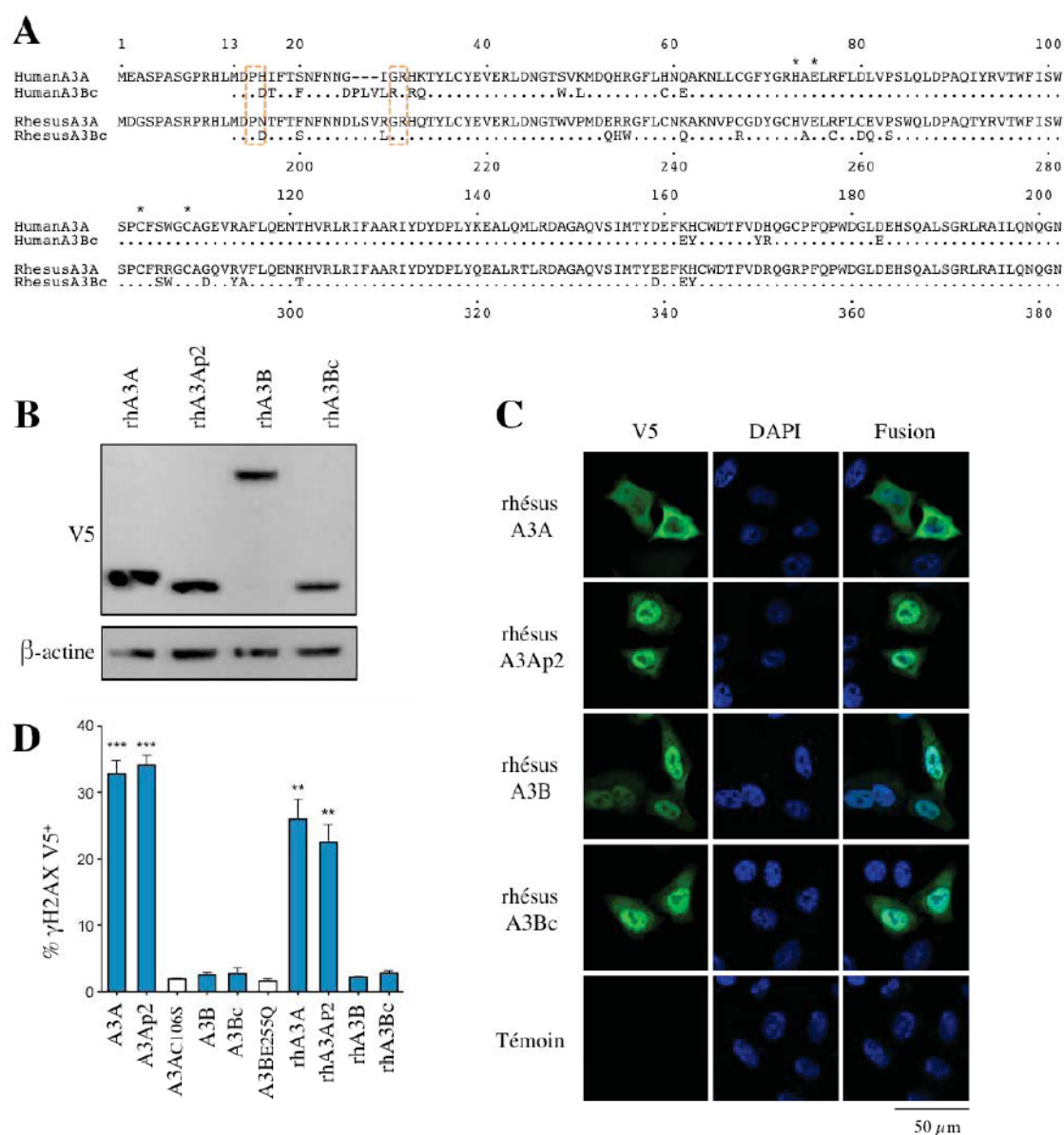


Figure 34 : Alignement des séquences d'A3A et A3Bc d'Homme et de macaque rhesus et localisation cellulaire des A3 de macaque rhesus.

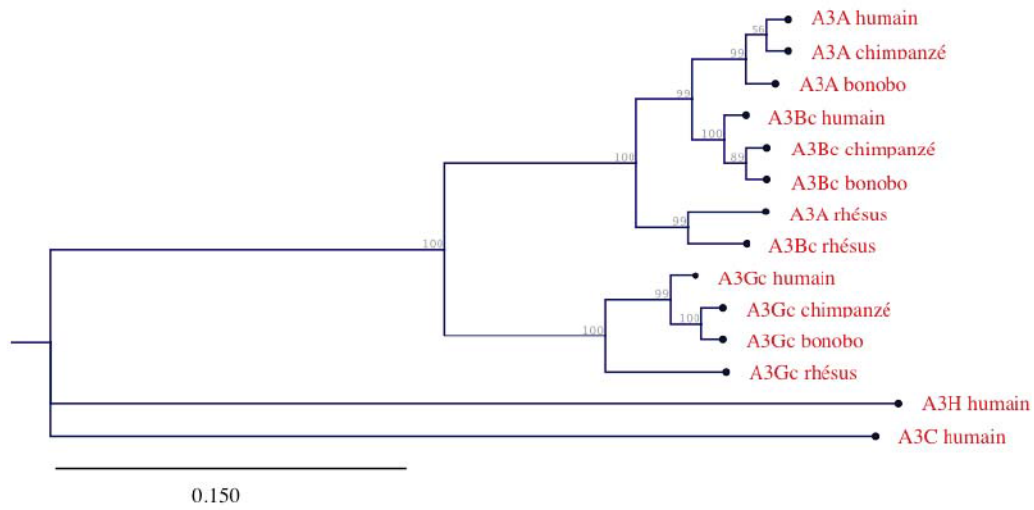
A) Alignement CLUSTALW des A3A et A3Bc d'Homme et de singe rhesus. Les principaux résidus impliqués dans la fonction enzymatique distinguant A3A et A3Bc chez l'Homme sont encadrés en orange, tandis que les résidus impliqués dans le domaine de coordination de zinc sont représentés par un astérisque. La numérotation du haut correspond à celle de l'A3A humaine tandis que celle du bas correspond à l'A3B de singe rhesus. B) Analyse par Western blot de l'expression des A3 de rhesus taggées V5 transfectées dans la lignée HEK293T. La β -actine sert ici de témoins de charge. C) Localisation cellulaire des protéines Rhésus A3. Les A3 taguées V-5 sont observées au microscope confocal dans la lignée HeLa 24h post transfection. Les noyaux sont marqués au DAPI. D) Analyse par FACS des cellules HeLa γ H2AX positives parmi les cellules V-5 positives 48h post transfection des mutants A3B. Les barres d'erreur sont calculées à partir de six transfections indépendantes. Les tests de significativité sont calculés par rapport aux mesures réalisées avec le mutant catalytique A3A_{C106S} en utilisant le test de Mann-Whitney (* $P < 0,1$, ** $P < 0,01$, *** $P < 0,001$).

matures de trois donneurs (Figure 36). Nous avons normalisé le nombre de copies des *APOBEC* par rapport au gène *PRMI* qui code la protamine, protéine spécifique du sperme. La fréquence des transcrits d'*APOBEC* observée est d'environ 4 à $16 \cdot 10^{-5}$ des lectures totales. *APOBEC4*, codant la protéine A4 appartenant à la famille des APOBEC mais dépourvue d'activité cytidine désaminase, est retrouvée à des niveaux un peu plus hauts (3 à $9 \cdot 10^{-4}$ des lectures totale) mais très inférieurs à ceux de *PRMI*, contrastant ainsi avec l'abondance d'A4 dans les testicules chez l'Homme (Rogozin et al., 2005). Sachant qu'environ 60% des spermatozoïdes présentent des problèmes de motilité, le faible nombre de transcrits d'APOBEC retrouvés en plus de la forte variabilité entre les échantillons, indique qu'il est très peu probable que les A3 aient un impact génotoxique sur le sperme.

C. Conclusion

Bien qu'A3A et A3B soient toutes deux des mutagènes chez l'Homme, ces deux enzymes ont un phénotype différent qui ne se recoupe pas. En effet, seule A3A induit des DSBs malgré le fait qu'elle soit moins retrouvée dans le noyau qu'A3B. Les mécanismes de cette dichotomie de génotoxicité sont multiples. D'une part, la région N-terminal d'A3B, A3Bn, joue un rôle important dans cette atténuation puisque l'introduction des résidus d'A3A dans les régions 1, 2 et 3 d'A3Bn ne suffit pas à restaurer l'induction de DSBs. De plus, la fusion d'A3Bn avec A3A diminue la formation de DSBs. Enfin, la substitution de résidus d'A3Bc par des résidus issus de la séquence d'A3A permet de restaurer partiellement l'activité DSB, indiquant une dégénérescence de cette région afin d'atténuer la génotoxicité d'A3B. De façon intéressante, cette dichotomie est aussi retrouvée chez le macaque rhésus. Néanmoins, le seul résidu correspondant à l'une des mutations différenciant A3A et A3B chez l'Homme est un polymorphisme chez le macaque rhésus sans impact sur la fonction de ces enzymes. Etant donné que les séquences d'A3A et d'A3B forment un groupe phylogénétique différent de celui formé par les séquences de l'Homme, du chimpanzé et du bonobo, il est probable que lors d'une conversion génétique plus récente, d'autres résidus aient subi une sélection ayant pour but le maintien de la faible génotoxicité d'A3B. Enfin, les A3 étant faiblement exprimées dans le sperme mature, aucune des A3 ne semble jouer un rôle dans la diversité génétique chez l'Homme.

A



B

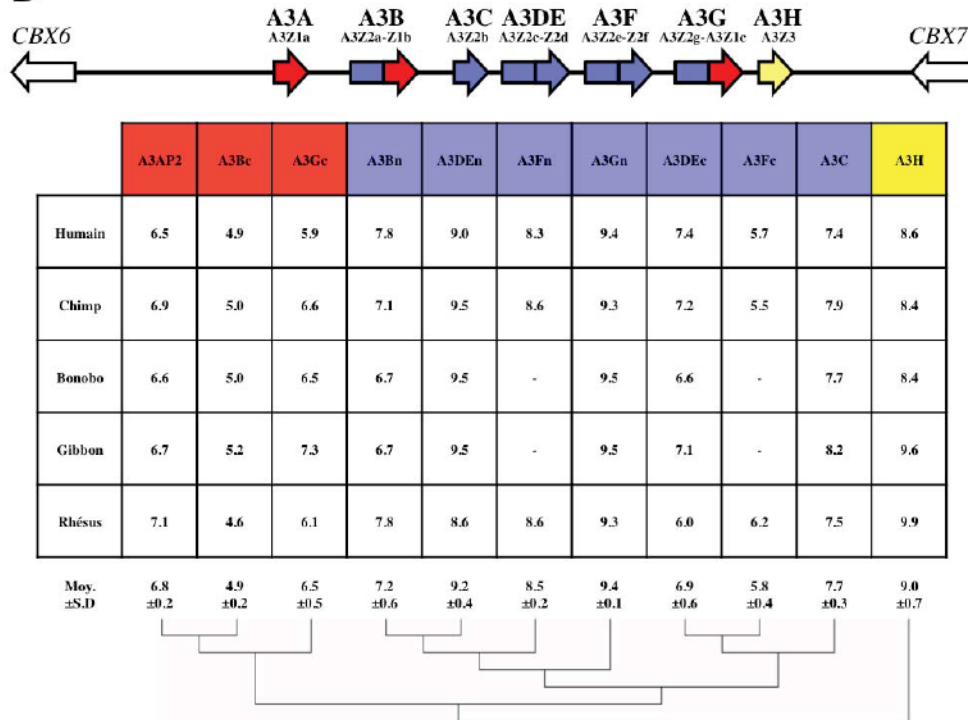


Figure 35 : Evolution du domaine Z1 d'A3A et d'A3B et point Isoélectrique théorique des différents domaines des A3 de primate.

Les nombres correspondant aux valeurs de *bootstrap* sont déterminés à partir de 1000 répliques. Les distances évolutives ont été calculées par la méthode de correction de Poisson et sont exprimées en nombre de substitutions par site. B) Point Isoélectrique (pI) théorique déterminé pour les domaines d'A3 de primate à partir de la séquence primaire d'acide aminés. Les valeurs ont été arrondies à la première décimale. Les domaines A3 sont organisés selon la phylogénie des domaines d'A3 humains.

Gene ID	Echantillon 1	Echantillon 2	Echantillon 3
AID	0	0	1
APOBEC1	0	0	1
APOBEC2	1	2	2
APOBEC3A	0	6	0
APOBEC3B	0	0	0
APOBEC3C	0	8	1
APOBEC3DE	0	6	0
APOBEC3F	0	8	2
APOBEC3G	0	11	2
APOBEC3H	0	5	0
APOBEC4	18	19	24
PRM1	26539	69783	27855
HPRT1	0	0	0

Figure 36 : Transcriptome des cytidines désaminases humaines dans le sperme mature humain purifié. Nombre absolu de transcrits RNAseq identifiés. PRM1 est le gène protamine 1, spécifique au sperme tandis que HPRT1 est décrit comme non exprimé dans le sperme.

III. Résultat 3 : Identification erronée d'ADN chromosomique édité par les APOBEC-3 dans les génomes de cancers.

A. Introduction

Les récents progrès dans le séquençage et l'étude des mutations génomiques dans les cancers ont permis de mettre en évidence le nombre important de mutations dans les cancers. Notamment, il a été mis en évidence que les mutations GC->AT représentent une proportion importante des mutations génomiques dans les cancers (Greenman et al., 2007). Dans certains génomes, on retrouve les mutations C->T majoritairement dans le contexte dinucléotidique 5'TpC (Alexandrov, Nik-Zainal, Wedge, Campbell, & Stratton, 2013; Nik-Zainal et al., 2012; Stephens et al., 2012). Fortes de cette observation, certaines équipes se sont intéressées au rôle que peuvent avoir les A3 dans la survenue des cancers, étant donné qu'on y trouve des mutations portant la signature des A3 (Burns, Lackey, et al., 2013; Shinohara et al., 2012). L'ADN hypermuté par les A3 peut être spécifiquement amplifié par la technique de 3DPCR (Suspène, Henry, et al., 2005). Cette technique, initialement utilisée pour la détection de génomes viraux hypermutés (Suspène, Aynaud, Koch, et al., 2011; Suspène, Guétard, et al., 2005; Vartanian et al., 2008), est de plus en plus employée pour amplifier de l'ADN nucléaire faiblement édité (Burns, Lackey, et al., 2013; Shinohara et al., 2012). C'est le cas pour deux équipes étudiant le rôle des A3 dans la survenue des cancers.

Ceux-ci ont employé la 3DPCR sur des cibles portant 2 à 13 mutations GC->AT par kilobases. Ils utilisent pour cela un gradient de température de dénaturation de 4-5°C sur un bloc de 12 puits. Ceci donne une différence de température entre les puits très faible, sachant que la température d'un puits peut varier de $\pm 4^\circ\text{C}$, laissant ainsi une place importante à la variabilité expérimentale. L'une des équipes présente les séquences suffisamment en détail : celles-ci ne sont pas mutées de façon monotone puisque qu'on retrouve des mutations C->T ou G->A sur le même brin d'ADN. De plus, le contexte dominant est le contexte 5'GpC, typique d'AID. Bien qu'AID puisse être impliquée dans les cancers, le faible taux de mutation peut aussi être le produit d'erreurs PCR. Dans la seconde étude, les auteurs attribuent à A3B un rôle dans l'*editing* de l'ADN génomique avec un signal 2 à 3 fois plus fort que le bruit de fond 3DPCR. Ce résultat nous a semblé particulièrement faible, notamment parce qu'A3B

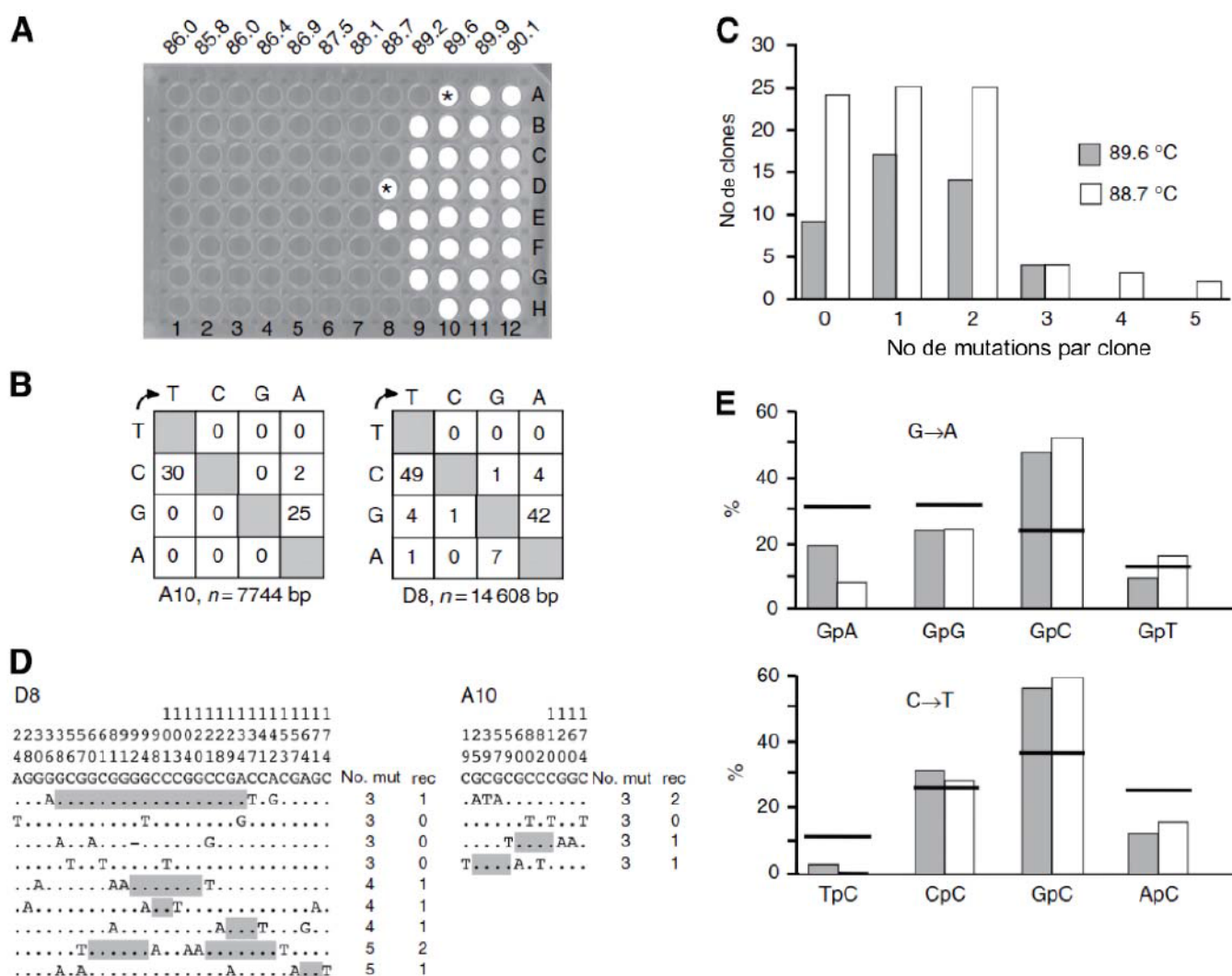


Figure 37 : Variation du bruit de fond selon la position sur le bloc lors d'une 3DPCR avec la *Taq* polymérase.

A) Les points blancs indiquent la présence de produits d'amplification pour la cible *TP53* de 176 pnb. Les astérisques indiquent les échantillons clonés et séquencés.

B) Matrices de mutation pour les séquences issues de A10 et D8. Les transitions sont systématiquement de type N->T, A ; n indique le nombre de bases séquencées.

C) Distribution du nombre de transitions CG->TA par clones. D) Sélection des séquences les plus mutées. Dans un souci de clarté, seuls les sites variables sont montrés, leur position étant indiquée au-dessus. La position doit être lue du haut vers le bas. Le nombre de mutations par clones (mut) et le nombre minimal de recombinaisons (rec) sont indiqués à droite. Les zones de recombinaison sont surlignées en gris. E) Contexte dinucléotidique 5' pour les transitions G->A et C->T. Les barres horizontales indiquent la valeur attendue.

mute très fortement l'ADN du VHB, à plus de 100 mutations par kilobases. Nous nous sommes donc intéressés au taux d'erreur de la 3DPCR en utilisant un clone ADN.

B. Résultats

1. Détermination du taux d'erreur lors d'une 3DPCR

Nous avons étudié le taux d'erreur de la 3DPCR à partir d'une région de *TP53* de 235 pb issue de la première étape de PCR. Cette région a été clonée dans un plasmide et nous servira à réaliser les différents 3DPCR. Cette région de la *TP53* ne possède aucune mutation et sert donc de référence. Nous avons ensuite réalisé une 3DPCR avec un gradient de température de dénaturation de 4°C. Nous observons que la température minimale de dénaturation varie de 0,9°C : en effet elle est de 89,6°C pour les lignes A et H et de 88,7°C pour les lignes du milieu D et E (Figure 37). Nous avons cloné et séquencé deux produits d'amplification, marqués par un astérisque. La majorité des séquences (70-80%) porte aussi bien des transitions C->T que G->A, jusqu'à 5 par clones, avec une moyenne de 7,5 transitions par kilobases, pour l'échantillon D8 (88,7°C) par comparaison avec la séquence initiale (Figures 37B et C). La séquence la plus divergente est présentée sur la Figure 37D. Pour l'échantillon A10 (89,6°C), le taux global de mutations est le même (7,4 par kb ; Figure 37B). Tandis que les transversions représentent environ 8% du nombre total de mutations, nous observons un biais en faveur des mutations N->A, T. Un tel phénomène est attendu étant donné que la 3DPCR amplifie de façon sélective les séquences AT riches et contraste fortement avec les mutations dues à la *Taq* polymérase qui génère plus de mutations A->G et T->C (ref) (Eckert & Kunkel, 1991).

Bien que la majorité des séquences porte aussi bien des mutations C->T et G->A, quelques séquences portent seulement un type de transition (Figure 37D). De façon intéressante, le contexte dinucléotidique des mutations dans ces 3DPCR est 5'GpC pour les deux types de transitions (Figure 37E). Un biais plus faible est observé pour le contexte 3', qui montre des mutations dans un contexte CpR où R représente un G ou un A (Figure 38A). Etant donné que les contextes 5' et 3' sont les mêmes pour les mutations G->A et C->T, on peut supposer que le même mécanisme mutationnel est responsable des mutations dans le motif 5'GpCpR.

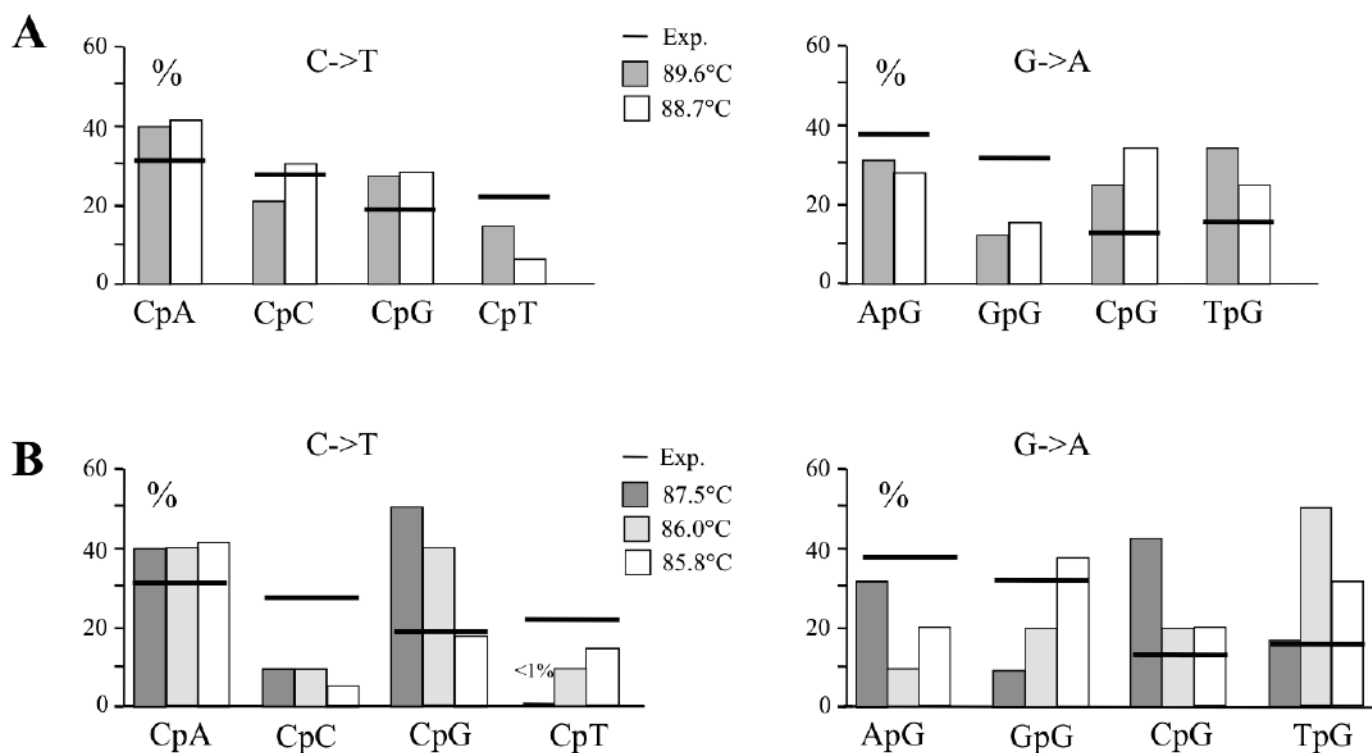


Figure 38 : Contexte dinucléotidique 3' des mutations bruit de fond de la 3DPCR avec un tampon standard ou *long range* et une *Taq* polymérase.

A) Fréquence des transitions C->T et G->A par contexte dinucléotidique 3' pour les échantillons A10 (89.6°C) et D8 (88.7°C) amplifiés avec un tampon standard et la *Taq* polymérase (voir figure 1). Les valeurs attendues sont indiquées par la barre. B) Fréquence des transitions C->T et G->A par contexte dinucléotidique 3' pour les échantillons A6 (87.5°C), B3 (86.0°C) et C2 (85.8°C) amplifiés avec un tampon *long range* et la *Taq* polymérase (voir figure 39).

Par ailleurs, la mixité des mutations peut être attribuée aux problèmes de recombinaison lors de la PCR (ref).

2. Détermination du taux d'erreur lors d'une 3DPCR avec un tampon *long range*

Dans des conditions acides, des sites abasiques peuvent apparaître sur l'ADN, entraînant l'insertion de A à ces endroits lors de la réplication. C'est entre autre pour éviter ce phénomène que sont utilisés les tampons de PCR *long range*, dont le pKa varie peu avec la température contrairement aux tampons standard Tris-HCl (Eckert & Kunkel, 1993). Nous avons par conséquent répété l'expérience précédente en utilisant la même enzyme mais dans un tampon *long range*. Sur le même gradient de 4°C, nous observons une variation de 1,7°C de la température minimale de dénaturation selon la position de l'échantillon sur le bloc chauffant (Figure 39A).

Encore une fois, des séquences portant aussi bien des transitions G->A que C->T (Figure 39B) avec la même préférence pour le contexte 5'GpC (Figure 39C et 39D) et un plus faible biais pour le contexte CpR en 3'(Figure 38B). De façon intéressante, le taux de mutation est environ deux fois plus important (4 à 20 mutations par kb) avec un tampon *long range* qu'avec un tampon standard (Figure 39B vs Figure 37B). Lorsque la *Pfu* est employée à la place de la *Taq*, des résultats similaires sont obtenus (séquences mixtes, contexte 5'GpC fréquence d'environ 13 mutations par kilobase ; Figure 40).

3. La 3DPCR permet de mettre en évidence des séquences faiblement éditées par les A3

Sur la base de ces observations, nous nous sommes proposés d'utiliser la 3DPCR sur de l'ADN de tissus cirrhotiques péri-tumoraux prélevés de patients souffrant de carcinomes hépatocellulaires (CHC) associés au VHB et au VHC de détecter des séquences faiblement éditées. Nous avons montré précédemment que de nombreuses A3 sont surexprimées dans le foie (ref) (Vartanian et al., 2010). Etant donné que le gène codant la β -caténine, *CTNNB1*, porte des mutations sur les résidus 32 à 35 dans environ 10 à 40% des CHC (ref) (Datta et al., 2008; Miyoshi et al., 1998), nous nous sommes concentrés sur un segment de 155 pb

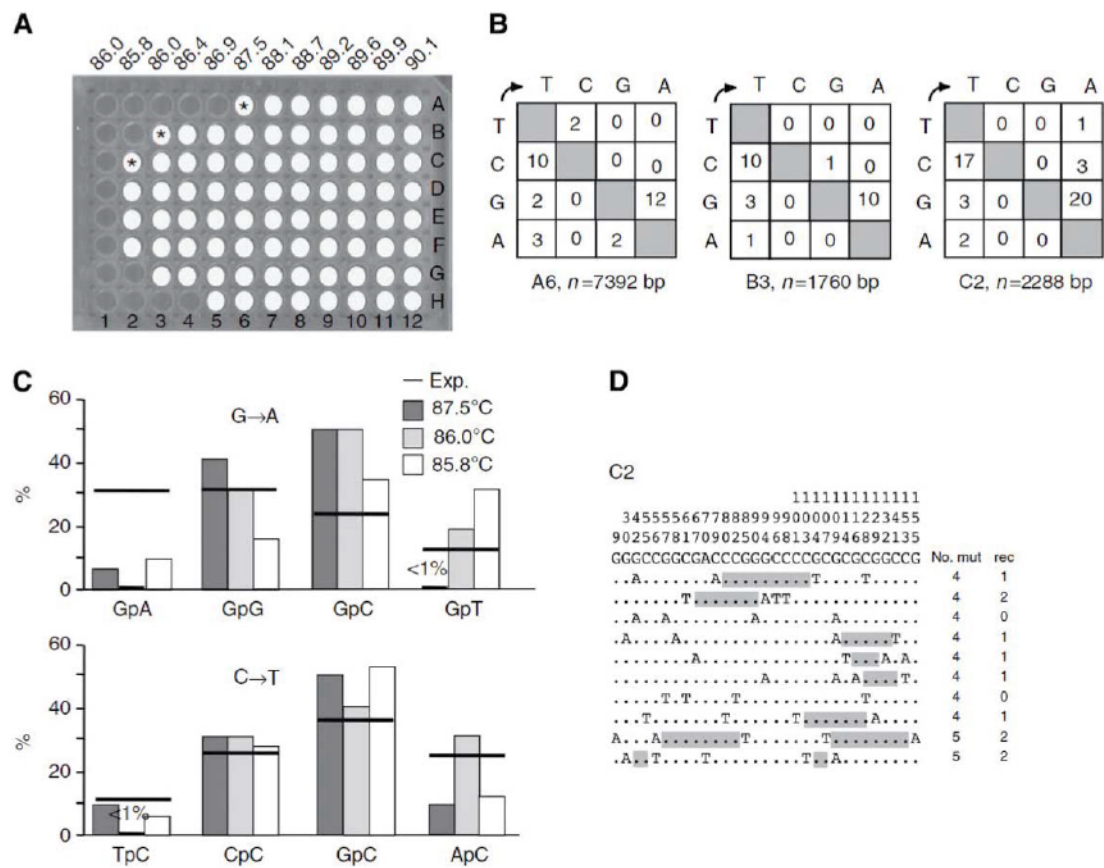


Figure 39 : Variation du bruit de fond selon la position sur le bloc lors d'une 3DPCR avec la *Taq* polymérase et le tampon *long range*.

A) Les points blancs indiquent la présence de produits d'amplification pour la cible *TP53* de 176 pnb. Les astérisques indiquent les échantillons clonés et séquencés. B) Matrices de mutation pour les séquences issues de A6, B3 et C2. Les transitions sont systématiquement de type N→T, A ; n indique le nombre de bases séquencées. C) Distribution du nombre de transitions CG→TA par clones. D) Sélection des séquences les plus mutées. Dans un souci de clarté, seuls les sites variables sont montrés, leur position étant indiquée au-dessus. La position doit être lue du haut vers le bas. Le nombre de mutations par clones (mut) et le nombre minimal de recombinaisons (rec) sont indiqués à droite. Les zones de recombinaison sont surlignées en gris.

encodant cette région. Trois échantillons cirrhotiques (n°310, 326 et 345) ont été analysés ainsi qu'un échantillon issu d'un foie saint (HL2). Les échantillons n°310 et 345 portent des mutations sur *CTNNB1* sur les tissus tumoraux tandis que ce n'est pas le cas pour l'échantillon n°326. Nous avons réalisé une 3DPCR en utilisant un gradient de dénaturation de 7°C (80-87°C) et une *Taq* polymérase. Nous avons ensuite cloné et séquencé le dernier point amplifié.

Encore une fois, les séquences portent aussi bien des mutations C->T que G->A et jusqu'à quatre substitutions par séquence. Seules quelques séquences ne portent qu'un type de substitution (Figures 41A et B). Le taux de mutation global pour l'échantillon n°310 est trois fois plus important que pour les échantillons n°326, 345 et HL2 et treize fois plus important que pour le plasmide contrôle (Figure 41C). Encore une fois, le contexte 5'GpC est prédominant (Figure 41D), ce qui exclue l'implication des A3 dans ces mutations.

A la lumière de ces résultats, nous avons voulu déterminer si il est possible de distinguer des séquences faiblement éditées par les A3 du bruit de fond 3DPCR décrit plus haut, et à quoi ressemblerait une séquence β -caténine éditée par les A3. Etant donné que les cellules hématopoïétiques expriment les A3 à des taux plus hauts comparés aux autres lignées cellulaires (<http://www.biogps.com>), nous avons analysé l'ADN extrait de lymphocytes T CD4⁺ purifiés à partir de PBMC et traités par de la PHA, de l'IL-2 et de l'interféron- α , ce qui a pour effet de fortement stimuler l'expression d'A3A (ref) (Aynaud et al., 2012). Afin d'améliorer la détection de séquences d'ADN nucléaire éditées par A3A, nous avons transduit à l'aide d'un lentivirus un plasmide pour l'UGI, un inhibiteur de l'UNG (Z. Wang & Mosbaugh, 1988). Nous avons ensuite amplifié la même séquence de la β -caténine par 3DPCR en utilisant un gradient de 7°C. Nous observons une différence de température minimale de dénaturation entre les cellules stimulées et non stimulées (contrôle) d'un degré Celsius (Figure 42A et C). Le clonage et le séquençage des séquences montrent que celles-ci portent des mutations monotones C->T ou G->A (Figures 4A à C) avec un fort biais pour le contexte TpC (GpA), ce qui est cohérent avec l'activité d'A3A et d'A3B (figures 42E et F).

Le témoin non stimulé ici nous apporte aussi des éléments intéressants. Celui-ci a été amplifié au centre du bloc, afin de minimiser les effets de l'hétérogénéité thermique et est amplifiable jusqu'à une température de dénaturation de 83,7°C (Figures 42A et B). Les séquences obtenues peuvent être classées en deux groupes de séquences AT riches (Figure 42C). Le premier groupe, composé de dix séquences, correspond à ce que nous définissons

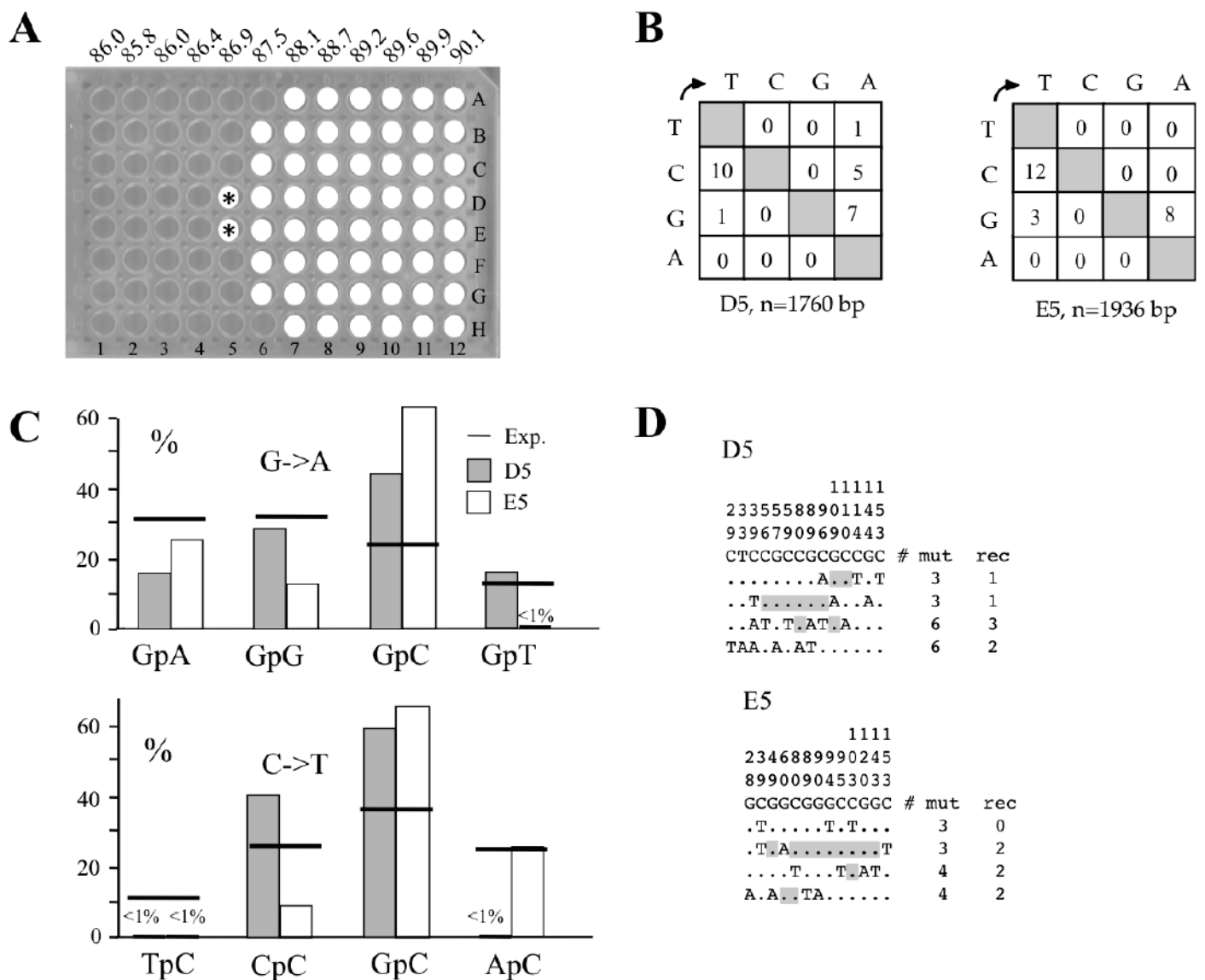


Figure 40 : Variation du bruit de fond selon la position sur le bloc lors d'une 3DPCR avec la *Pfu* polymérase

A) Les points blancs indiquent la présence de produits d'amplification pour la cible *TP53* de 176 pnb. Les astérisques indiquent les échantillons clonés et séquencés. B) Matrices de mutation pour les séquences issues de D5 et E5. Les transitions sont systématiquement de type N->T, A ; n indique le nombre de bases séquencées. C) Contexte dinucléotidique 5' pour les transitions G->A et C->T. Les barres horizontales indiquent la valeur attendue.

D) Sélection des séquences les plus mutées. Dans un souci de clarté, seuls les sites variables sont montrés, leur position étant indiquée au-dessus. La position doit être lue du haut vers le bas. Le nombre de mutations par clones (mut) et le nombre minimal de recombinaisons (rec) sont indiqués à droite. Les zones de recombinaison sont surlignées en gris.

comme le bruit de fond de la 3DPCR et dont les séquences portent jusqu'à cinq mutations par locus (32 mutations par kilobase), impliquant autant de transitions C->T que G->A avec un biais pour les transitions dans le contexte GpC. Le second groupe de séquences présenté dans la figure 42C est caractérisé par des mutations monotones G->A, au nombre de cinq ou six, dans un contexte 5'TpC et 5'CpC. En effet, le contexte dinucléotidique est ici similaire à celui obtenu avec la β -caténine ainsi qu'avec des cellules stimulées par l'interféron- α sur la région CTNNB1. Ainsi, ceci démontre que la 3DPCR permet d'amplifier et de mettre en évidence des séquences faiblement éditées par les A3.

La constante présence de séquences portant des transitions C->T et G->A, quelle que soit l'origine de l'ADN amplifié (clone moléculaire, extrait de tissus ou issu de PBMC) est intrigante. Nous avons précédemment montré que les recombinaisons lors de la PCR contribuent à la complexité de l'ADN amplifié (Meyerhans, Vartanian, & Wain-Hobson, 1990). Ceci a souvent lieu à la fin de l'amplification, lorsque la concentration d'ADN est supérieure à celle de polymérase, empêchant celle-ci de finir la synthèse des brins en cours d'élongation avant le changement de température accompagnant le passage à l'étape de dénaturation. Des temps d'élongation plus long réduisent le taux de recombinaisons (Meyerhans et al., 1990). Lors des expériences précédentes impliquant *TP53* et *CTNNB1*, nous avons fixé le temps d'élongation à 10 minutes. Néanmoins, la 3DPCR a pour particularité d'avoir des températures de dénaturation proches de la température d'élongation. Certaines matrices pourraient ne pas être complètement dénaturées, entraînant ainsi un blocage de l'élongation par la polymérase et rendant l'amplification moins efficace. C'est pourquoi nous réalisons les 3DPCR en 42 cycles alors que 30 cycles suffisent à amplifier la cible TP53 lors d'une PCR classique.

4. La 3DPCR induit une complexification des séquences mutées par recombinaison

Afin de considérer la question de la recombinaison lors de la PCR, nous avons mélangé 1 ng d'ADN de trois clones moléculaires de *CTNNB1* afin de réaliser une 3DPCR avec le même gradient de 7°C. Nous avons ensuite cloné et séquencé le dernier point positif d'amplification (84,7°C). En comparant la séquence initiale au produit de PCR (figure 42D), nous pouvons observer que la recombinaison par PCR entraîne une complexification de la

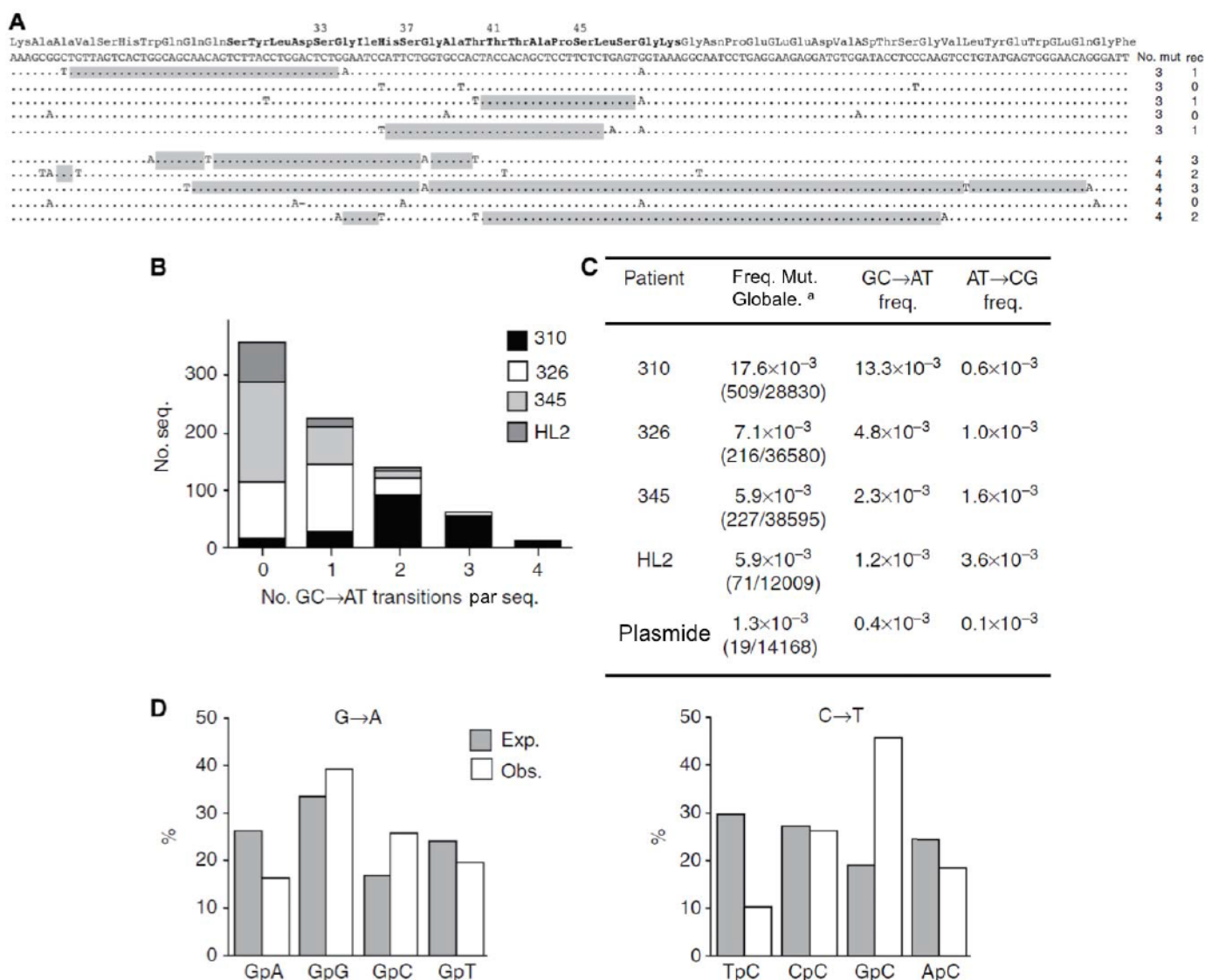


Figure 41 : 3DPCR sur des produits PCR de *CTNNB1* issus de foie sains ou cirrhotique
A) Sélection de produits de 3DPCR sur *CTNNB1* provenant de cirrroses virales. La séquence en amino-acides est indiquée au-dessus des séquences nucléotidiques. Les résidus fréquemment substitués dans les cancers du foie sont indiqués en gras. Le reste des annotations est similaire à celles de la figure 1D. B) Histogramme empilé du nombre de transitions GC→AT par séquence par échantillon de foie. C) Fréquence de mutation par échantillon avec le nombre absolu de mutations et le nombre de bases mutées et séquencées (entre parenthèse). D) Contexte dinucléotidique 5' des transitions GC→AT pour les transitions G→A et C→T respectivement.

séquence. L'hypothèse la plus parcimonieuse pour expliquer la présence de séquences portant des mutations C->T et G->A est que la polymérase produit un large spectre de mutations. Parmi les mutations CG->TA, il va y avoir dominance de mutations C->T dans le contexte GpCpR ou G->A dans le contexte YpGpC. Les cycles successifs de 3DPCR vont entraîner un enrichissement en ADN portant des transitions CG->TA et, dans une moindre mesure, des transversions N->TA qui sont sélectivement amplifiées. La complexité des séquences est apportée par la recombinaison lors de la PCR.

C. Conclusions

La technique de 3DPCR a initialement été employée dans le but d'amplifier des plasmides et des génomes viraux édités (Suspène, Aynaud, Koch, et al., 2011; Suspène, Guétard, et al., 2005; Vartanian et al., 2008). Cette technique a ensuite permis de mettre en évidence de l'ADN mitochondrial et nucléaire hyperédité. (Suspène, Aynaud, Guétard, et al., 2011). Deux études plus récentes ont mis en évidence par 3DPCR des séquences d'ADN nucléaire édité par les A3 (Burns, Lackey, et al., 2013; Shinohara et al., 2012). Néanmoins, le taux de mutation observé dans ces études est proche du taux d'erreur que nous avons déterminé dans cette étude. Dans l'un des cas, la différence de température de dénaturation entre les échantillons positifs et le contrôle négatif ne dépasse jamais 1°C tandis que la fréquence de mutations est de 6 à 9 mutations par kb, seulement deux fois plus élevé que le taux d'erreur basal (Shinohara et al., 2012). De plus, les séquences portent des mutations mixtes C->T et G->A avec pour contexte préférentiel le contexte 5'GpC pour les mutations C->T. Ces observations ne sont pas sans rappeler les résultats obtenus dans les figures 37 à 41 et suggèrent des mutations non physiologiques générées par la 3DPCR.

Par ailleurs, on peut supposer que les séquences mixtes sont les résultats d'un *editing* par les A3, dans un mécanisme où un brin édité serait répliqué et dont le brin néo-synthétisé serait alors à nouveau édité. Néanmoins, le contexte dinucléotidique observé doit être celui des A3 dans ce cas et non pas celui du bruit de fond de la 3DPCR comme dans les études citées précédemment. A ce jour, seules les séquences hypermutées du VHB possèdent des mutations mixtes C->T et G->A (Suspène, Guétard, et al., 2005). De plus, le taux de mutation est bien supérieur à ce qui est observé ici. Enfin, ces mutations respectent le contexte dinucléotidique des A3, quel que soit le brin considéré.

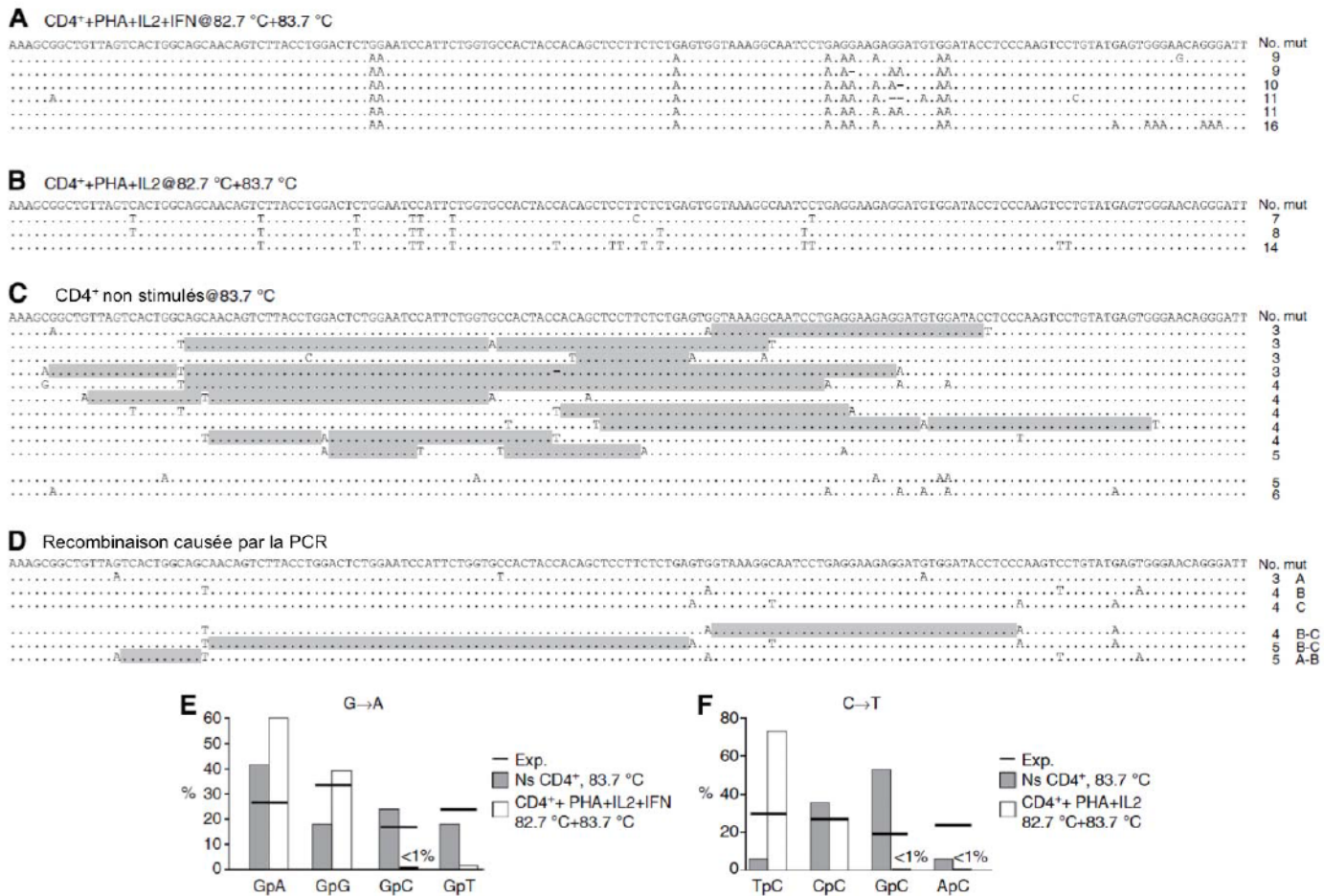


Figure 42 : Séquences *CTNNB1* hypermutées par les A3 purifiées de lymphocytes T CD4 humains.

A) Sélection de séquences issues de cellules CD4⁺ traitées à l'interféron α , portant la marque de l'editing par les A3. Ces séquences portent des mutations monotones avec un biais pour le contexte GpA. B) Sélection de séquences issues de cellules CD4⁺ humaines, qui cette fois sont hypermutées C->T. C) Sélection de séquences amplifiées à partir de cellules CD4⁺ non stimulées servant de contrôle pour les expériences précédentes. Le premier groupe de 10 séquences présente des séquences portant des mutations mixtes C->T et G->A. La figure est annotée comme la Figure 1D. Le second groupe de deux séquences porte des substitutions G->A de façon monotone avec un biais pour le contexte GpA. D) Trois clones moléculaires de *CTNNB1* mutés (A, B et C) ont été mélangés et amplifiés par 3DPCR. Les trois séquences du dessous sont des recombinants des séquences initiales et appelées B et C et A et B. Les régions de recombinaison sont surlignées en gris. E) et F) Contexte dinucléotidique 5' des transitions G-> et C->. Les valeurs attendues sont indiquées par la barre horizontale. NS : non stimulé.

La technique de 3DPCR ne permet pas d'identifier les transitions C->T fixées dans les génomes de cellules tumorales. En effet, la fréquence globale de mutation dans ces génomes est d'environ 10^4 à 10^5 mutations par génomes, soit 0.003 à 0.03 par kb. Ceci est bien inférieur au taux d'erreur de la 3DPCR qui est d'environ 4 à 20 mutations par kb, soit 12 à 60 millions de mutations par génomes diploïdes. Etant donné le nombre de mutations détectées et le contexte 5'GpC associé à celles-ci, ces mutations sont vraisemblablement le produit de la 3DPCR.

La 3DPCR constitue donc un outil puissant pour la détection de séquences hyperéditées par les A3, mais ne permet pas de conclure quant aux séquences hypoéditées dont la fréquence de mutation est inférieure à 20 mutations par kb. Dans ce cas, il est nécessaire de cloner et de séquencer les produits d'amplification, afin de déterminer la nature des mutations (mixtes ou monotones) et surtout de vérifier que le contexte dinucléotidique correspond à celui des A3.

B

	# cellules analysées (85°C)		
	Total	Mutées	f %
P2	512	03	0.6
CD4 ⁺ Donneur 1	192	33	17.2
CD4 ⁺ Donneur 2	190	23	12.0
P2+act.D	217	51	24.6
P2+etop.	288	64	22.2

B

Séquences analysées (95°C)

Cell	#seq ^a	1≤ Hypomutation (CG->TA) ≤5			Hypermutation ≥6		
		# hypo ^b	% hypo ^c	f _{CG->TA} ^d	# hyper ^e	#mut ^f	
P2	B02	82	42	51	2.4 10 ⁻³	3	10, 18, 28
	E11	90	48	53	2.7 10 ⁻³	3	11, 18, 26
CD4 ⁺ Donneur 1	A06	128	67	52	6.1 10 ⁻³	2	23, 36
	F09	125	63	50	6.0 10 ⁻³	2	18, 30
CD4 ⁺ Donneur 2	E02	117	59	50	5.9 10 ⁻³	3	06, 26, 35
	C08	94	49	52	6.7 10 ⁻³	2	34, 36
P2+act.D	A04	116	60	52	4.9 10 ⁻³	3	06, 12, 14
	F06	134	71	53	4.8 10 ⁻³	2	09, 43
P2+etop.	D08	89	48	54	5.3 10 ⁻³	0	0
	E10	110	58	53	5.2 10 ⁻³	1	11

C

[illegible]

Figure 43 : Editing massif de l'ADNcymt par les A3 dans une cellule.

A) Fréquence de cellules uniques comportant de l'ADNcymt édité par les A3 obtenue par PCR avec une Td de 85°C. B) Nombre de séquences d'ADNmt hypo- ou hypermutées obtenues par PCR à Td de 95°C ; ^a nombre de séquences analysées ; ^b nombre de séquences hypomutées, ^c proportion de séquences hypomutées par cellule ; ^d moyenne de la fréquence de mutations CG->TA dans les séquences hypoéditées, ^e nombre de séquences d'ADNmt identifiées ; ^f Nombre (#) de transitions C->T ou G->A par séquence d'ADNmt. C) Sélection de séquences MT-COI hypermutées. Bien que l'editing ait lieu sur les deux brins, seul le brin codant (positif) est représenté.

IV. Résultat 4 : L'induction de la surexpression de la cytidine désaminase APOBEC3A par l'ADN cytoplasmique endogène provoque des dommages de l'ADN chromosomique.

A. Introduction

Les A3 ont été initialement décrites en tant que facteurs de restriction viraux. Il a été montré que celles-ci sont actives sur de nombreux génomes viraux parmi lesquels des rétrovirus (VIH) ou des virus à ADN (HPV) (Lecossier et al., 2003; Mariani et al., 2003; Turelli, Mangeat, Jost, Vianin, & Trono, 2004; Vartanian et al., 2010). Ce mécanisme de défense se base en partie sur la réponse interféron de la cellule hôte et la surexpression d'A3A et d'A3G qui en résulte (Argyris et al., 2007; Bonvin et al., 2006; Koning et al., 2009). Par ailleurs, il a été montré que ces A3 sont actives sur l'ADN mitochondrial et l'ADN chromosomique étant donné que ceux-ci portent les marques de l'activité des A3, c'est-à-dire des mutations de type C->T dans le contexte dinucléotidique 5'TpC(Suspène, Aynaud, Guétard, et al., 2011).

Il est maintenant connu que la réponse immunitaire et l'inflammation peuvent être induites par la détection d'ADN cytoplasmique par l'intermédiaire de nombreux systèmes de détection tels que les récepteurs TLR ou la protéine RIG-I(Ablasser et al., 2009; Y. H. Chiu, Macmillan, & Chen, 2009; Fernandes-Alnemri, Yu, Datta, Wu, & Alnemri, 2009; West et al., 2015 ; P. Yang et al., 2010).

A la lumière de nombreux travaux démontrant que l'inflammation chronique peut être la source de nombreux cancers (Balkwill & Mantovani, 2001), nous nous sommes demandé dans quelle mesure l'ADN cytoplasmique peut entraîner la surexpression des A3, et notamment d'A3A, par l'intermédiaire des censeurs cytoplasmiques de l'ADN.

Pour cela, nous avons analysé la fréquence d'*editing* de l'ADN mitochondrial au niveau d'une cellule. Ainsi, nous observons que 25% des cellules étudiées présentent des milliers de génomes mitochondriaux hyperédités. Nous observons par ailleurs que, dans la lignée monocytaires humaine THP-1, la transfection d'ADN double brin aboutit à la transcription de celui-ci par l'ARN polymérase 3 conduisant à un intermédiaire ARN double brin reconnu par RIG-I. Cette interaction conduit à la production d'IFN-I et à l'induction

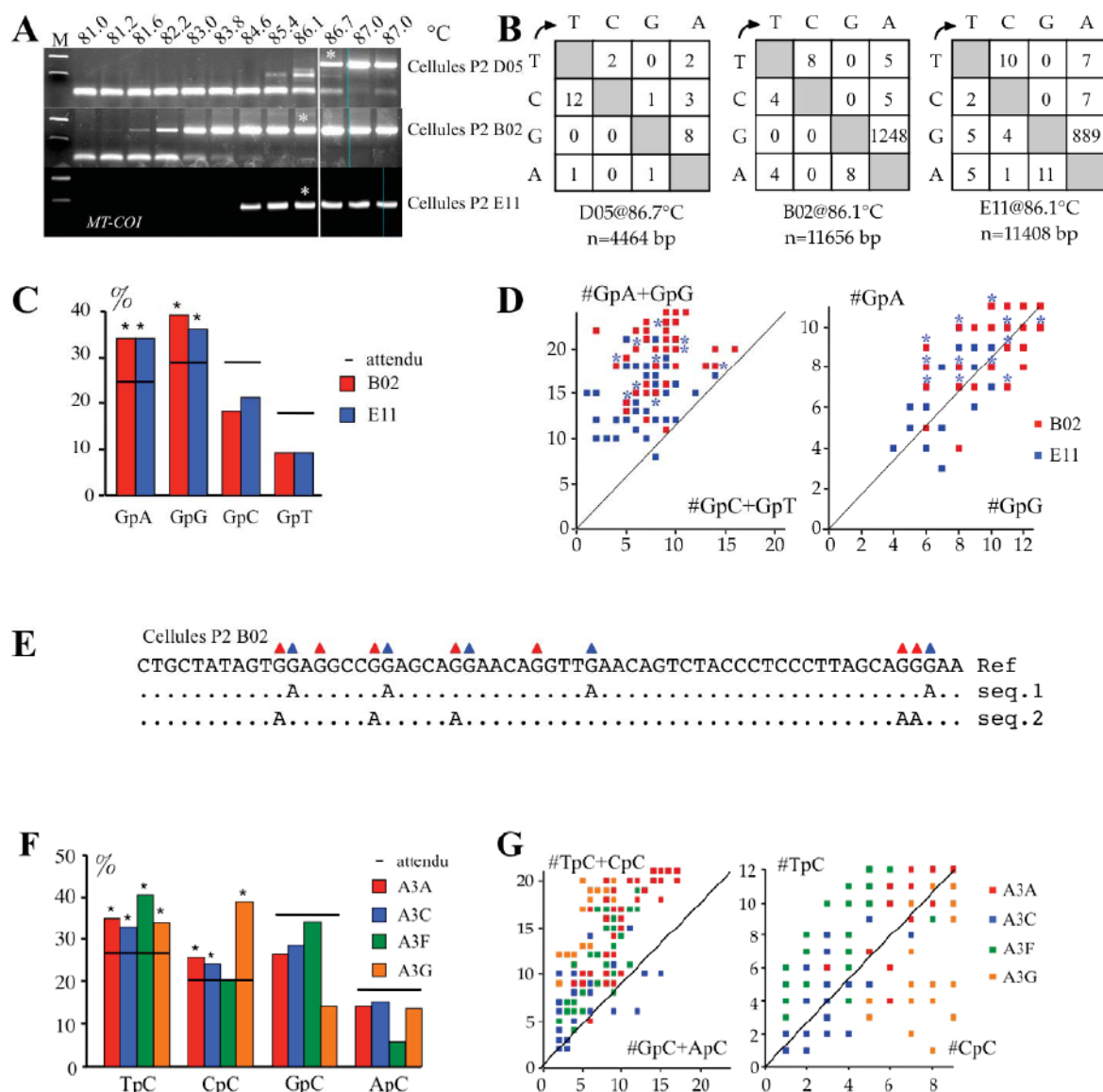


Figure 44 : Editing de la séquence *MT-COI* dans une cellule P2

A) Amplification de l'ADN *MT-COI* jusqu'à une Td 86.1°C pour la cellule B02 et 84,6°C pour la cellule E11. La ligne blanche indique le seuil délimitant les amplicons édité et non édité. La cellule D05, qui ne porte pas d'ADNmt édité, est utilisée comme control. M : marqueur de poids moléculaire. B) Matrice des mutations des séquences hyperéditées de *MT-COI* des cellules D05, B02 et E11. n indique le nombre de bases séquencées. C) Contexte dinucléotidique des mutations sur le brin négatif de l'ADN dans la région *MT-COI* des cellules B02 et E11. La significativité a été déterminée par un test χ^2 (* $p < 0,05$). D) Analyse par clonage de l'editing de séquences *MT-COI* individuelles mutées des cellules B02 et E11. L'astérisque bleu représente les séquences se recoupant entre B02 et E11. Le nombre de cibles TpC+CpC vs. GpC+ApC éditées par séquence est représenté respectivement en x et y (droite) et l'analyse des clones sur TpC vs. CpC (gauche). E) Editing de l'ADN *MT-COI* des cellules P2 et B02 au niveau d'une cellule dans deux contextes nucléotidiques GpA (clone 1) et GpG (clone 2). F) Contexte dinucléotidique de l'ADN *MT-COI* de cellules de cailles édité par les cytidines désaminases A3A, A3C, A3F et A3G (* $p < 0,05$). G) Analyse de l'editing des clones de *MT-COI* pour A3A, A3C, A3F et A3G.

d'A3A, qui entraîne la désamination et le catabolisme de l'ADN transfecté. Nous montrons ici que les cytidines désaminase et notamment A3A ont un rôle anti-inflammatoire et qu'il existe une balance entre catabolisme et mutation de l'ADN génomique, qui sont des événements « accidentels ».

B. Résultats

1. Editing de l'ADN mitochondrial cytoplasmique (ADNcymt) au niveau d'une cellule

L'ADN cytoplasmique mitochondrial est l'une des cibles les plus sensibles pour la détection de l'*editing* par les A3. Nous nous sommes d'abord intéressés aux cellules P2, une lignée de cellules B déficientes pour l'UNG et immortalisées par l'EBV, qui constitue un modèle de choix car celle-ci permet une détection de l'activité des A3 plus aisée (Suspène, Aynaud, Guétard, et al., 2011). Nous avons trié et isolé par trieur de cellules des cellules P2 viables sur plaques PCR 96 puits. Nous avons ensuite procédé à une PCR puis à une 3DPCR sur le gène mitochondrial *MT-COI* sur plus de 500 cellules afin de détecter les mutations par les A3. Sur 512 cellules, seulement trois (0,6%) sont éditées au niveau de la mitochondrie (Figure 43A). Les premiers « run » de PCR des cellules B02, E11 et de la cellule négative D05 ont été soumis à une 3DPCR en utilisant un gradient de température. Comme nous pouvons l'observer sur la Figure 44A, la cible peut être amplifiée jusqu'à des températures restrictives de 81,6°C et 84,6°C pour les cellules B02 et E11. En revanche, il n'y a pas d'amplification d'ADN mitochondrial en dessous de 86,7°C, température minimale de dénaturation à laquelle l'ADN mitochondrial normal est amplifié dans les cellules D05. Les clonages et séquençage des produits de 3DPCR marqués d'un astérisque (Figure 44A) montrent un *editing* important du brin antisens du gène *MT-COI*, environ 40 à 50 fois au-dessus du bruit de fond (cellule D05, Figure 44B). Le contexte dinucléotidique indique un fort biais pour les contextes TpC (ou GpA sur le brin complémentaire) et CpC (ou GpG) (Figures 44C et 44D) ce qui correspond au contexte de désamination des A3. L'analyse au niveau d'une région de *MT-COI* montre que la majorité des 5'TpC et 5'CpC est éditée, avec certains

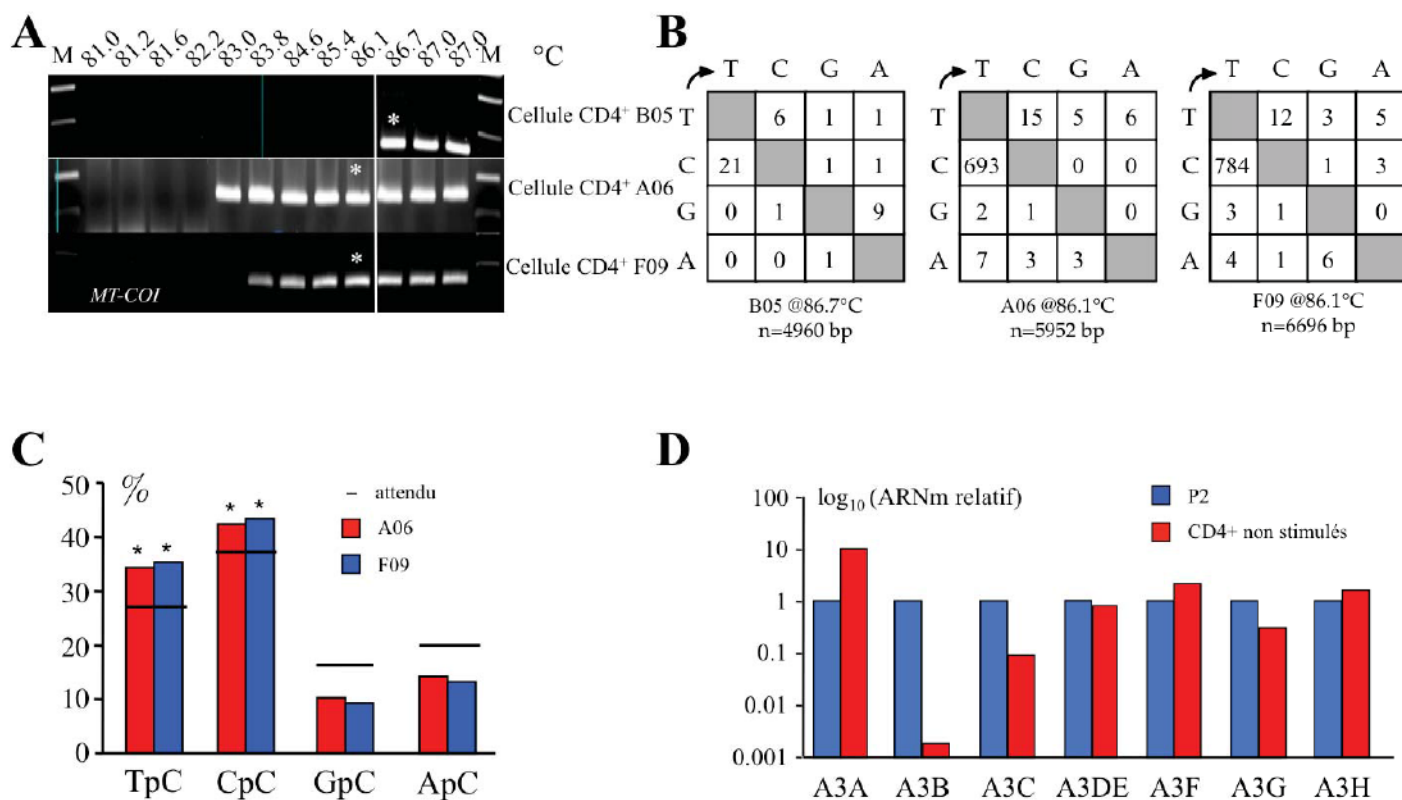


Figure 45: Editing de *MT-COI* dans les lymphocytes T CD4+ provenant du donneur 1.

A) Amplification par 3DPCR de l'ADN *MT-COI* édité jusqu'à 83-83,8°C pour les cellules A06 et F09. La ligne blanche indique le seuil entre les amplicons édités et non édités en terme de température de dénaturation. La cellule B05 qui ne présente pas de signe d'editing sert de contrôle. L'astérisque indique les séquences clonées et séquencées. M : Marqueur de poids moléculaire. B) Matrices de mutation pour les séquences de l'ADN *MT-COI* hyperditées des cellules A06 et F09 (Td=86,1°C). n indique le nombre de bases séquencées. C) Contexte dinucléotidique du brin négatif de la région *MT-COI* obtenu des cellules A06 et F09 à Td=86,1°C. Le test χ^2 est employé afin de déterminer les contextes dinucléotidiques dont la fréquence diffère significativement des autres (* $p < 0,05$). D) Analyse par TaqMan de cellules T CD4+ sélectionnées positivement du donneur 1. Les données, en duplicate, ont été normalisées par rapport à *RPL13A*. Le tout a été normalisé par rapport aux cellules P2 non traitées afin de faciliter la comparaison.

sites au sein desquels seules les bases 5'TpC ou 5'CpC sont éditées, conformément au degré de processivité typique de ces enzymes (Figure 44E).

Afin de déterminer l'enzyme A3 impliquée dans ces mutations, la lignée QT6 a été transfectée par des plasmides exprimant différentes A3 et l'*editing* de l'ADNmt a été analysé. La lignée QT6 n'exprimant pas d'A3, l'emploi de cette lignée permet de s'affranchir du bruit de fond. On observe que A3A, A3C, A3F et A3G induisent toutes des mutations sur l'ADNmt, comme cela a déjà été montré au cours de travaux précédents (Suspène, Aynaud, Guétard, et al., 2011). La comparaison entre le contexte obtenu ici et celui obtenu sur les cellules P2 (Figure 44F) ou lors de l'analyse sur une molécule indique qu'A3A et A3C sont les candidats les plus probables (Figure 44G). L'absence d'*editing* sur l'ADNmt de caille par A3B peut s'expliquer par une translocation très rapide de cette protéine dans le noyau.

2. Editing de l'ADNcymt dans les lymphocytes T CD4⁺ primaires

De nombreuses études ont déjà mis en évidence l'expression des A3 dans les leucocytes (Aynaud et al., 2012; Koning et al., 2009). Afin d'avoir un modèle plus physiologique, nous avons purifié les lymphocytes T CD4⁺ de deux donneurs sains (D1 et D2). Les cellules ont ensuite été triées par tri cellulaire et analysées de la même façon que les cellules P2 précédemment (Figure 43A). On observe par PCR avec une Td de 85°C qu'environ 17% et 12% des cellules D1 et D2 (respectivement) sont mutées sur le gène *MT-COI* (Figure 43A), soit 20 à 28 fois plus que ce qui est observé dans les cellules P2. Deux cellules isolées de D1 (A06 et F09, Figure 45A) et de D2 (C08 et E02, Figure 46A) ont été analysées en détail. LA 3DPCR permet l'amplification de la cible bien au-delà de la température de référence de 86,7°C (figures 45A et 46A). De plus, l'analyse des mutations (Figures 45B et 46B) et l'analyse des séquences mettent en évidence l'*hyperediting* par les enzymes A3, avec encore une fois un fort biais en faveur des contextes 5'TpC et 5'CpC, typiques des cytidines désaminases A3 (Figures 45C et 46C). Afin de mesurer l'expression relative des gènes *A3*, nous avons étudié le transcriptomes par TaqMan sur les sept A3, avec comme gène de référence *RPL13A*. Nous pouvons observer que les CD4⁺ expriment bien plus d'A3A et moins d'A3B que les cellules P2. Ceci suggère que A3A est responsable de la

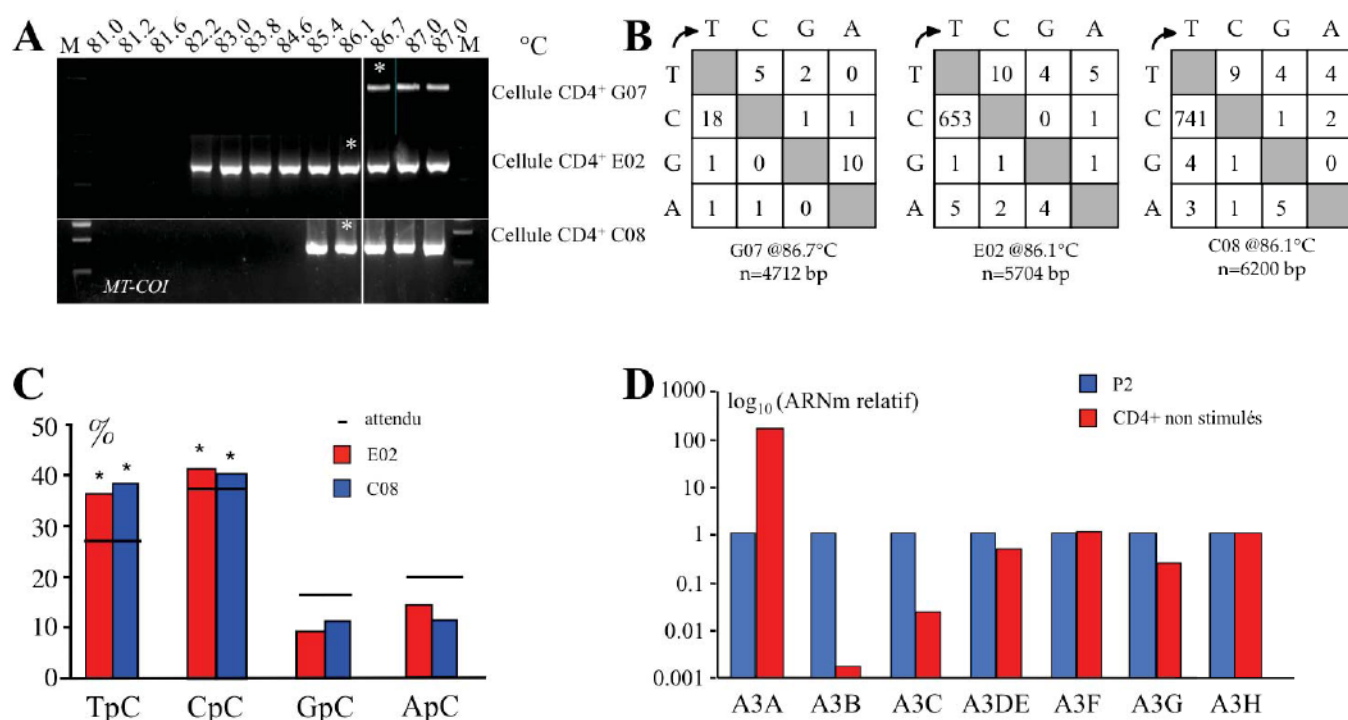


Figure 46 : Editing de MT-COI un lymphocyte T CD4⁺ provenant du donneur 2.

A) Amplification par 3DPCR de l'ADN *MT-COI* édité jusqu'à 82,2-85,4°C pour les cellules E02 et C08. La ligne blanche indique le seuil entre les amplicons édités et non édités en terme de température de dénaturation. Les cellules G07 qui ne présentent pas de signe d'editing servent de contrôle. L'astérisque indique les séquences clonées et séquencées. M : Marqueur de poids moléculaire. B) Matrices de mutation pour les séquences de l'ADN *MT-COI* hyperédité des cellules E02 et C08 (Td=86,1°C). n indique le nombre de bases séquencées. C) Contexte dinucléotidique du brin négatif de la région *MT-COI* obtenu des cellules E02 et C08 à Td=86,1°C. Le test χ^2 est employé afin de déterminer les contextes dinucléotidiques dont la fréquence diffère significativement des autres (* $p < 0,05$). D) Analyse par TaqMan de cellules T CD4⁺ du donneur 1 sélectionnées positivement. Les données, en duplicate, ont été normalisées par rapport à *RPL13A*. Le tout a été normalisé par rapport aux cellules P2 non traitées afin de faciliter la comparaison.

majorité des mutations dans les cellules primaires dont l'ADNmt est muté (Figures 45D et 46D).

3. L'etoposide et l'actinomycine D induisent l'apoptose et la surexpression des A3

Etant donné la faible fréquence de cellules P2 présentant de l'ADNcymt hyperédité, probablement lié à des niveaux relativement bas de transcrits d'A3A, nous avons testé l'impact des stress cytotoxiques sur la désamination de l'ADNcymt. L'etoposide est un inhibiteur de la topoisomérase II tandis que l'actinomycine D se lie à l'ADN et inhibe l'initiation de la transcription aboutissant à la formation de DSBs et à l'apoptose (Mussil et al., 2013). Le traitement des cellules P2 par ces molécules entraîne le relargage du cytochrome c et l'apoptose après 16h de traitement (Figures 47B et 47C). L'analyse du transcriptome montre une surexpression d'A3A, A3F, A3G et A3H lorsque les cellules sont traitées à l'etoposide et une surexpression d'A3A seulement après traitement à l'actinomycine D (Figure 47A).

Après un traitement des cellules P2 de 16 heures avec 100µM d'etoposide ou d'actinomycine D, nous avons trié les cellules par FACS dans des plaques 96 puits. Comme précédemment, nous avons étudié les hypermutations de l'ADNmt à une Td de 85°C. Nous observons que le traitement par ces drogues augmente d'environ 40 fois le nombre de cellules dont l'ADNmt est hyperédité (Figure 43A). Les produits d'amplification de la première PCR de deux cellules traitées par 100µM d'etoposide (D08 et E10) ou 100µM d'actinomycine D (A04 et F06) ont été analysés en détail comme décrit précédemment (Figure 47D) et comparés à une cellule non traitée (C08). Les produits d'amplifications obtenus à une Td de 86,1°C ont été clonés et séquencés. Nous observons que le brin antisens de l'ADNmt est très fortement hyperédité (Figure 47F), et que ces mutations sont préférentiellement situées dans le contexte 5'TpC et 5'CpC (Figure 47E). Ceci met en évidence l'implication des A3 dans le stress génotoxique résultant du traitement à l'etoposide et l'actinomycine D.

Etant donné que la 3DPCR favorise l'amplification de l'ADN hyperédité plutôt que l'ADN hypoédité, cette technique est peu quantitative (Suspène, Henry, et al., 2005). Afin

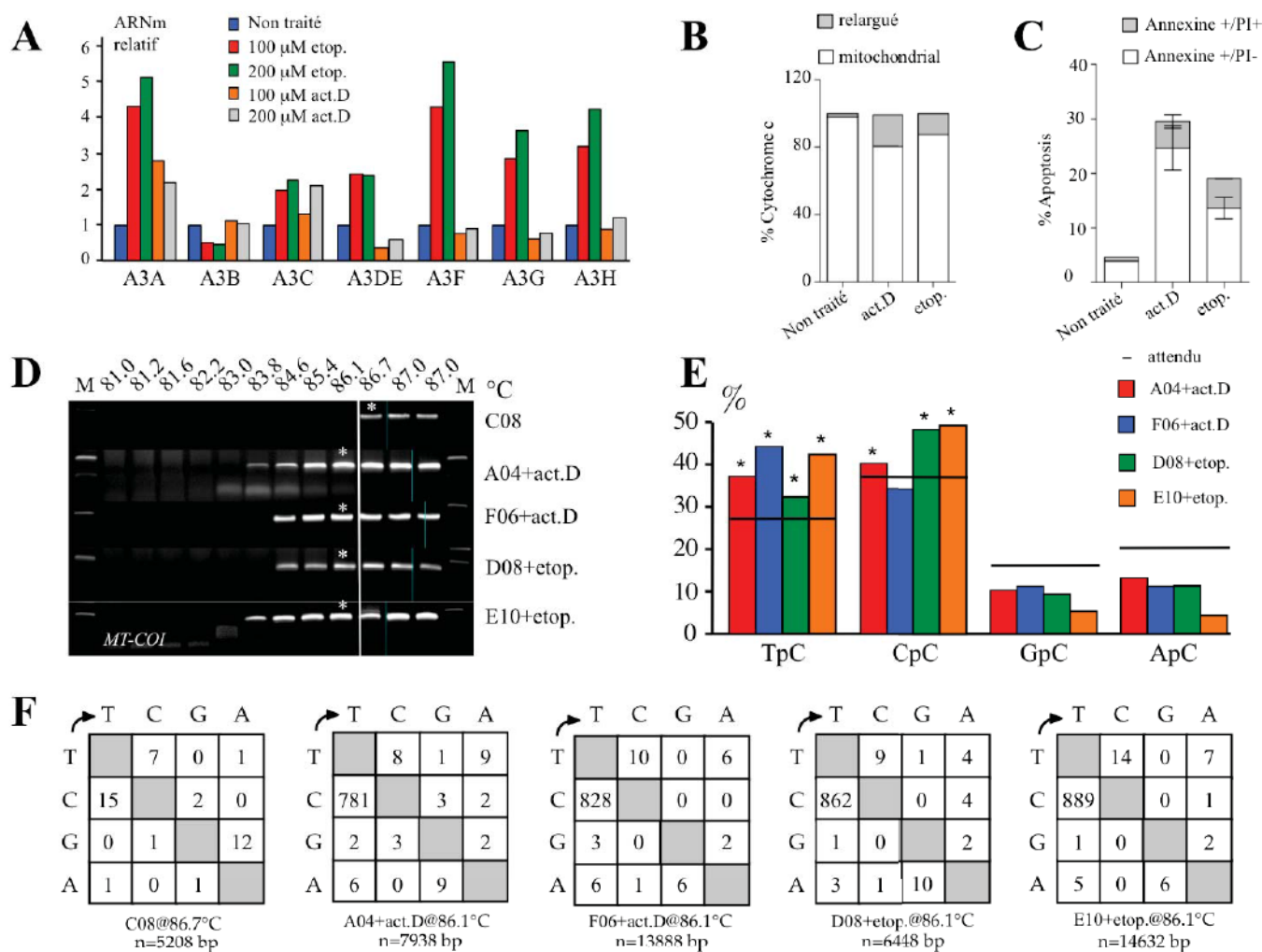
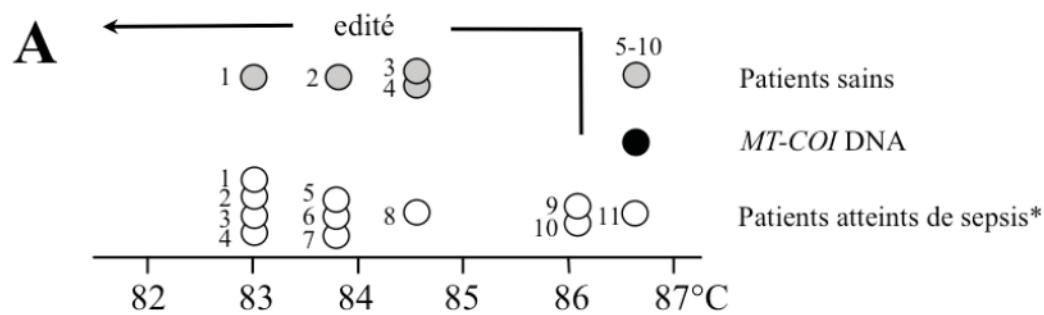


Figure 47 : L'étoposide et l'actinomycine D induisent l'expression des A3 et l'editing de l'ADN MT-COI dans les cellules P2

A) Profil d'expression des A3 dans les cellules P2 traitées avec de l'étoposide ou avec de l'actinomycine D. L'expérience et les mesures ont été réalisées en duplicate. Les niveaux d'expression sont rapportés à celui de *RPL13A*, notre gène de ménage, et aux cellules P2 non traitées afin de faciliter la comparaison. B) et C) Analyse par FACS du relargage de cytochrome c (A) et de l'apoptose (B) dans les cellules P2 après 16h de traitement avec 100µM d'actinomycine ou 100µM d'étoposide. Le marquage annexin V permet de rendre compte de l'apoptose précoce tandis que l'iodure de propidium (IP) indique l'apoptose et la nécrose. Les moyennes et les écarts types sont calculés sur deux expériences indépendantes. D) Amplification par 3DPCR de l'ADN *MT-COI* édité jusqu'à une Td de 83,8-84,6°C pour les cellules A04 et F06 traitées à l'actinomycine D et pour les cellules P2 D08 et E10 traitées à l'étoposide. La cellule C08 sert ici de control négatif. La ligne blanche indique le seuil entre les séquences éditées et non éditées en terme de température de dénaturation. Les astérisques indiquent les échantillons clonés et séquencés. M : Marqueur de poids moléculaire. E) Contexte dinucléotidique de l'editing par les A3 de l'ADN *MT-COI*. Le test Chi-2 a été utilisé pour déterminer la significativité de différence entre les valeurs (* $p < 0,05$). F) Matrice des mutations pour l'ADN *MT-COI* hyperédité dérivés des clonages d'amplicons de 3D-PCR. n indique le nombre de bases séquencées.



B

Patients	IL6 ^a	%CG->TA ^b	YpC/RpC ^c
S01	1973	54	3.23
S02	621	64	3.91
S03	15181	49	2.77
S04	2135	49	2.77
S05	2356	59	2.34
S06	3200	73	4.46
S07	144	59	3.39
S08	677	57	2.16
S09	2315	35	1.76
S10	14507	45	2.35
H01	nd	58	3.49
H02	nd	53	3.80
H03	nd	59	3.23
H04	nd	36	1.38

Figure 48 : Editing de *MT-COI* chez les patients sains et atteints de sepsis.

A) Représentation schématique des plus basses températures de dénaturation obtenues par 3DPCR pour l'ADN *MT-COI* amplifié à partir de sérum de patients sains ou atteints de sepsis. Le cercle noir représente la température minimale de dénaturation d'un clone moléculaire de *COI-MT*, tandis que les cercles blancs représentent les échantillons des patients atteints de sepsis et les cercles gris des patients sains (* Test de Fisher, $p=0,024$). B) Analyse des séquences hyperéditées obtenues des patients atteints de sepsis (S) et sains (H). ^a Concentration d'IL6 sérique (pg/mole), ^b pourcentage de mutations G->A, ^c Le ratio YpC/RpC à été calculé selon la formule suivante : $((TpC+CpC)/(GpC+ApC))_{\text{observé}}/((TpC+CpC)/(GpC+ApC))_{\text{attendu}}$. Une valeur supérieure à 1 indique une désamination par les A3 dans le contexte TpC+CpC. nd : non déterminé.

d'étudier l'ADNcymt peu édité, nous avons réalisé une PCR à une Td de 95°C sur deux cellules des échantillons identifiés dans la figure 47A. Nous avons ensuite cloné et séquencé les amplicons. Sachant que ces données concernent l'ADNmt localisé aussi bien dans la mitochondrie que dans le cytoplasme, il est intéressant de noter qu'environ 50% des séquences portent 1 à 5 mutations GC->AT tandis que 1-4% sont hypermutés dans la majorité des échantillons (Figure 47B) et présentent un contexte 5'TpC. On observe que les séquences hypermutées obtenues à 95°C sont fortement mutées sur les deux brins (Figure 47C). La présence de séquences hypomutées indique une sous-estimation de la proportion de cellules dont l'ADNcymt est hypermuté présentée dans la figure 43A. Ces résultats laissent supposer que la majorité des fragments d'ADNcymt amplifiables portent des mutations GC->AT.

4. ADNmt circulant dans les cas de sepsis.

Afin de faire le lien entre le stress cellulaire et l'ADNmt cytoplasmique, nous avons recherché un modèle plus physiologique. Le sepsis est caractérisé par une inflammation systémique causée par le développement invasif d'une bactérie dans la circulation sanguine (Schulte, Bernhagen, & Bucala, 2013). La même démarche de PCR/3DPCR que précédemment a été employée pour amplifier le gène *MT-COI* à partir d'ADN du sérum de onze patients atteints de sepsis (Fitting et al., 2012) comparé à l'ADN de dix donneurs sains (Figure 48A). La cible est amplifiée à des Td allant jusqu'à 83,1°C, bien moins que pour la cible non mutée (Td d'environ 86,5°C, Figure 48A), chez dix des onze des individus malades (environ 90%). En revanche, seuls quatre des dix sérums sains contrôles arborent des signes d'*editing* (Figure 48A). Le clonage et le séquençage confirment l'impact des A3 avec les biais classique 5'TpC et 5'CpC dans les contextes dinucléotidiques (Figure 48B). Aucune corrélation n'a été mise en évidence entre l'importance de l'*editing* et le niveau d'IL6 sérique. Etant donné que la proportion d'échantillons positifs pour l'*editing* de l'ADNmt est plus important chez les patients atteints de sepsis que chez les donneurs sains contrôles, ceci peut refléter un fort microenvironnement inflammatoire et le relargage de cellules nécrotiques dans la périphérie.

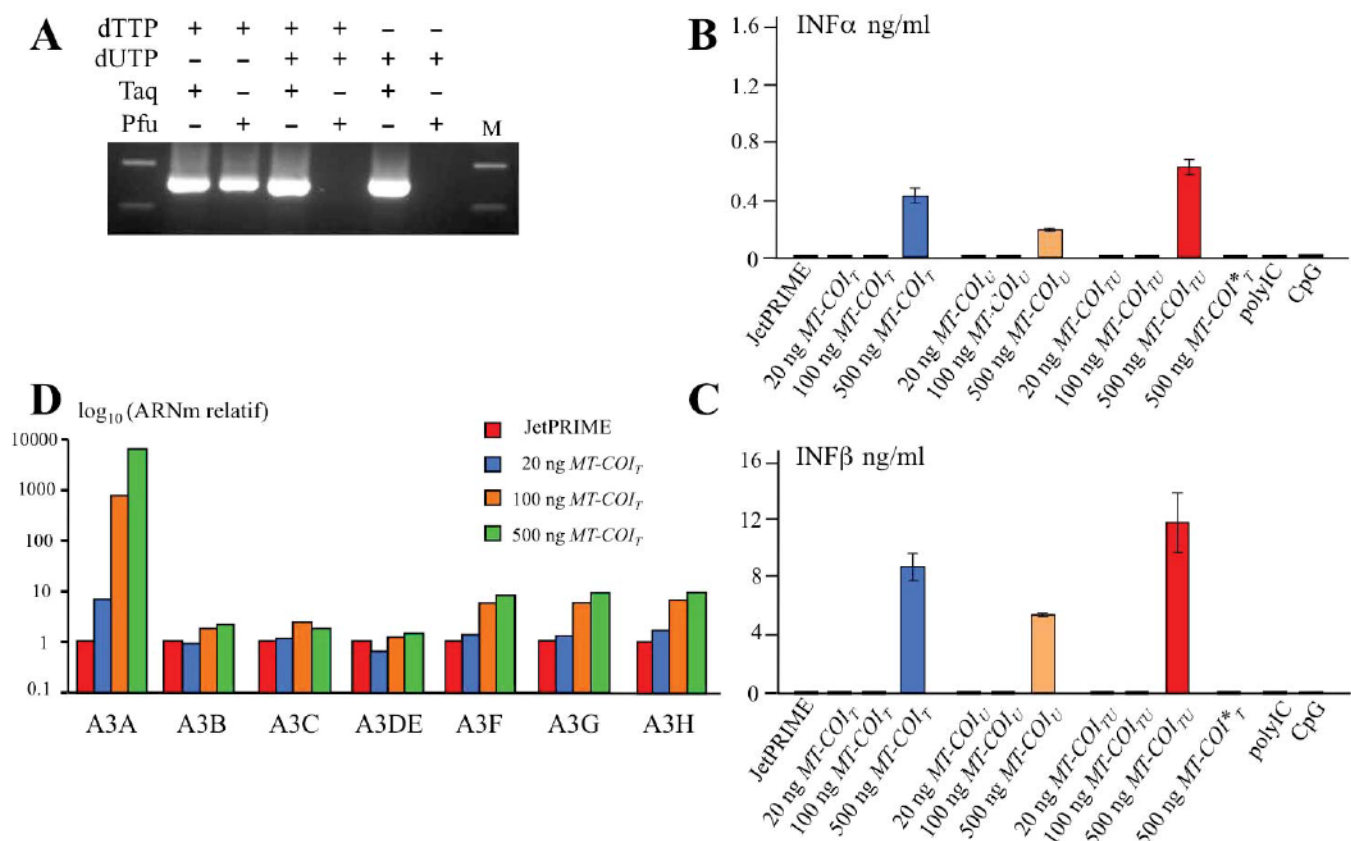


Figure 49 :L'ADN contenant de l'uracile ne constitue pas un nouveau signal de danger.

A) Migration sur gel d'agarose des produits d'amplification de la PCR *MT-COI* réalisée avec de la *Taq* ou de la *Pfu* polymérase en présence de dTTP, dUTP ou dTTP+dUTP. M, marqueur de poids moléculaire. B) Production d'interféron α et β après transfection des THP-1 par des fragments PCR d'ADN contenant des résidus dT ou dU. *MT-COI** indique une incubation avec de l'ADN sans transfection. D) Profil transcriptionnel des APOBEC3 dans les cellules THP-1 transfectées. Les données ont été normalisées par rapport au gène de référence RPL13A. Les mesures ont été réalisées en duplicate et normalisées par rapport aux cellules traitées au JetPRIME afin de faciliter la comparaison. E) 3DPCR sur l'ADN *MT-CYB* des cellules THP-1 transfectées. La ligne blanche délimite les amplicons édités des amplicons non édités. L'astérisque indique les échantillons clonés et séquencés. M : Marqueur de poids moléculaire. F) Fréquence de distribution des mutations G->A par clones en fonction de la quantité d'ADN transfectée.

5. L'Uracile dans l'ADN cytosolique ne constitue pas un signal de danger

Etant donné que l'ADNcymt hyperédité porte des résidus uridines, normalement absents de l'ADN, nous avons envisagé que la présence d'uridine dans l'ADN puisse constituer un signal de danger pour la cellule. Afin d'étudier cette hypothèse, le gène *MT-COI* a été amplifié en utilisant soit des dUTP, soit un mélange équimolaire de dTTP et dUTP ou seulement du dTTP. Afin de contrôler l'incorporation des dUTP dans l'ADN, une seconde PCR a été réalisée avec des amorces internes et en employant une polymérase *Pfu* ou *Taq* (Figure 49A). Comme toutes les ADN polymérases d'archée, la *Pfu* ne peut pas amplifier les fragments d'ADN portants des résidus dU. Comme attendu, la PCR *Pfu*, contrairement à la PCR *Taq* n'a pas permis d'amplifier l'ADN lorsque les résidus dUTP (*MT-COI_U*) ou le mélange dUTP+dTTP (*MT-COI_{UT}*) ont été incorporés (Figure 49A).

La lignée THP-1, issue d'une leucémie aigue lymphoblastique, est inductible par l'IFN α et est l'une des rares lignées dans laquelle toutes les A3A sont exprimées ou éventuellement inductibles, en particulier A3A. Nous avons transfecté la lignée THP-1 avec la matrice

MT-COI_T, *MT-COI_U* ou *MT-COI_{UT}*. Après 24h de transfection, nous avons clarifié les surnageants et mesuré la production d'IFN α et β par ELISA. Comme nous pouvons l'observer sur les figures 49B et 5C, 500ng de *MT-COI_T*, *MT-COI_U* ou *MT-COI_{UT}* induisent une forte réponse IFN comparé au contrôle négatif JetPrime, sans qu'il n'y ait de différences selon la composition de l'ADN. Ceci indique donc que la présence d'uridine dans l'ADN ne constitue pas un signal de danger pour la cellule. De plus, les THP-1 incubées, et non pas transfectées, avec 500 ng d'ADN *MT-COI_T* (*MT-COI_T**), ne produisent pas d'IFN α et β . Ce résultat montre que c'est l'ADN intracellulaire qui déclenche cette réponse (Figure 49B et 49C). Cette réponse ne passe pas par la voie des TLR3 et 9 étant donné que les THP-1 sont insensibles aux CpG non méthylés et aux polyIC (Zarembek & Godowski, 2002) (Figure 49B et 49C).

Le traitement des THP-1 par l'ADN *MT-COI_T* induit de manière dose-dépendante une surexpression d'A3A, A3F, A3G et A3H après 24h de transfection (Figure 49D). De façon intéressante, cette surexpression est observée même lorsque l'interféron n'est pas détectable dans le surnageant (Figure 49E). On observe qu'A3A est très fortement surexprimée lors de la transfection, faisant de cette A3 la plus sensible à la transfection d'ADN. La réponse

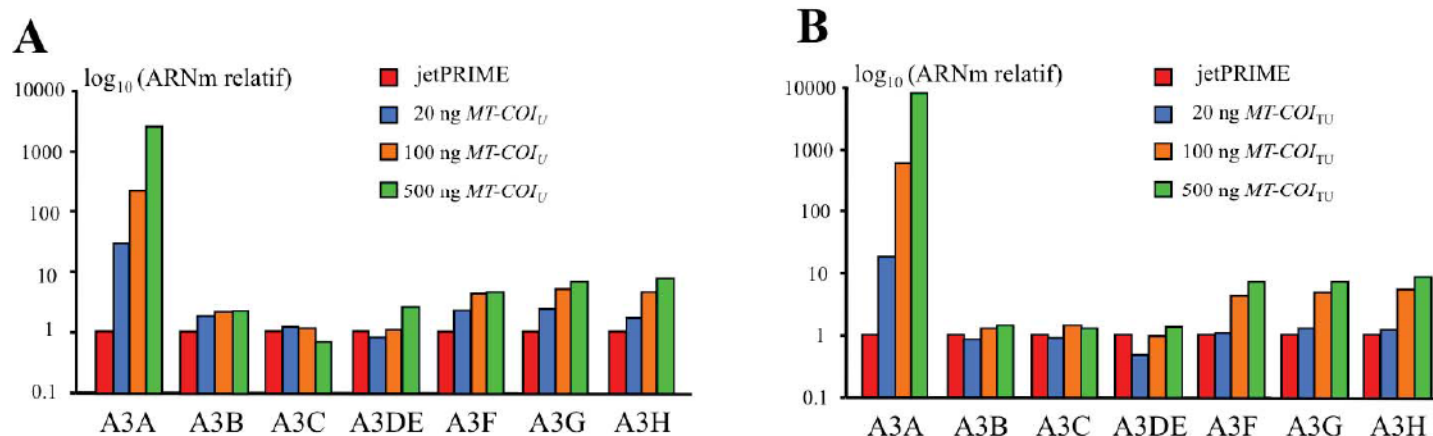


Figure 50 : Transcritome des A3 dans les cellules THP-1 transfectées par de l'ADN.
 Les transcriptomes ont été déterminés pour l'ADN contenant des dU (A) ou un mélange 50:50 de dT et dU (B). Les duplicats ont été normalisés par rapport au niveau du gène de ménage *RPL13A* et normalisés aux cellules THP-1 non traitées afin de faciliter les comparaisons.

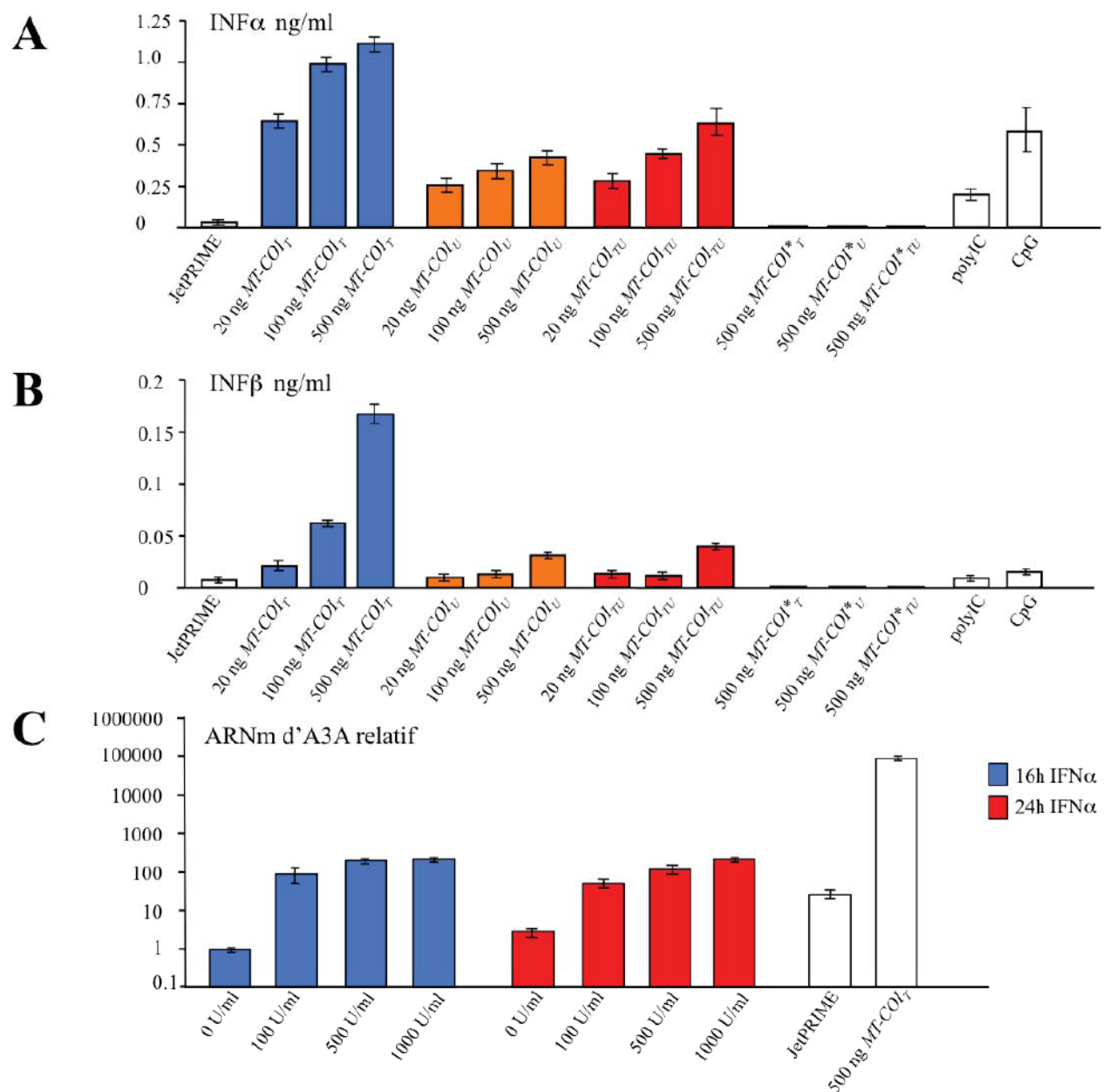


Figure 51 : La transfection d'ADN MT-COI active la production d'IFN α et β dans les PBMC.

Les productions d'interféron α (A) et d'interféron β (B) ont été mesurées après traitement avec de l'ADN polyIC (500ng/ml) et CpG (500ng/ml). MT-COI* indique une incubation avec de l'ADN sans transfection. C) Profil transcriptionnel d'A3 dans les THP1 traitées ou non traitées avec de l'IFN α 16h ou 24h. Des cellules transfectées avec 500ng de MT-COIT ou traitées au JetPRIME pour 24h servent ici de contrôle.

similaire des A3 quelle que soit la composition de l'ADN transfecté (*MT-COI_U*, *MT-COI_{UT}*, Figures 50A et 50B) renforce l'hypothèse que l'uracile dans l'ADN ne constitue pas un signal de danger pour la cellule, aboutissant à la production IFN α et à la surexpression d'A3A.

6. L'ADN *MT-COI* active la réponse interféron conduisant à l'expression des A3

Afin de voir si l'ADN exogène peut entraîner la surexpression des A3 et aboutir à l'*editing* de l'ADNcymt endogène, nous avons transfecté la lignée THP-1 avec l'ADN *MT-COI_T* et effectué une 3DPCR sur le cytochrome b de l'ADNmt (*MT-CYB*) (Suspène, Aynaud, Guétard, et al., 2011). Nous observons une diminution de la Td nécessaire lorsque les cellules sont transfectées (Figure 49E). Les amplicons de 3DPCR amplifiés à 80,4°C ont été clonés et séquencés. Nous observons que les séquences portent de mutations C->T, la fréquence moyenne d'*editing* augmentant avec la quantité d'ADN transfecté (Figure 49F).

Etant donné la très faible efficacité de transfection des cellules THP-1, moins d'un pourcent en utilisant un plasmide d'expression GFP comme marqueur, la surexpression d'A3A est vraisemblablement le résultat de la transfection de quelques cellules dont la réponse est amplifiée par l'effet paracrine d'IFN α et β via l'induction d'IFNAR. Forts de cette observation, nous avons transfecté des PBMC non stimulés, connus pour être difficile à transfecter de façon efficace, avec l'ADN *MT-COI_T*. Le niveau d'induction d'IFN α est comparable à celui obtenu avec la lignée THP-1 (Figure 49B et figure 51A), tandis que la production d'IFN β est environ 40 fois moins importante (*MT-COI*_{T, U, UT}*). L'importance de l'induction d'IFN indique une amplification vraisemblablement paracrine. En effet, le traitement des cellules THP-1 avec l'IFN α durant 16h ou 24h augmente l'expression d'A3A de façon dose-dépendante (Figure 51C).

Afin de déterminer si l'ADN simple brin ou double brin est responsable de la production d'IFN, quatre paires d'oligonucléotides complémentaires correspondant à de petites régions de MT-COI ont été synthétisées : WT1+2 (43 pb, 51% GC) et WT3+4 (39pb, 56%GC) tandis que Hyp1+2 (43pb, 7%GC) et Hyp3+4 (39pb, 13% GC) correspondent aux

Primer	Séquence
WT ₁	5' ACTCTTACCTCCCTCTCTCCTACTCCTGCTCGCATCTGCTATA
WT ₂	5' TATAGCAGATGCGAGCAGGAGTAGGAGAGAGGGAGGTAAGAGT
Hyp ₁	5' ATTTTATTTTTTTTTTTTTTTTATTTTGTGTTGTATTGTTATA
Hyp ₂	5' TATAACAAATACAAACAAAAATAAAAAAAAAAAAAATAAAAAAT
WT ₃	5' AACTACTCCCACCCTGGAGCCTCCGTAGACCTAACCATC
WT ₄	5' GATGGTTAGGTCTACGGAGGCTCCAGGGTGGGAGTAGTT
Hyp ₃	5' AATTATTTTTATTTTGGAGTTTTTGTAGATTTAATTATT
Hyp ₄	5' AATAATTAAATCTACAAAACTCCAAAATAAAAAATAGTT

Tableau 4: Compendium des oligonucléotides utilisés dans cette étude.

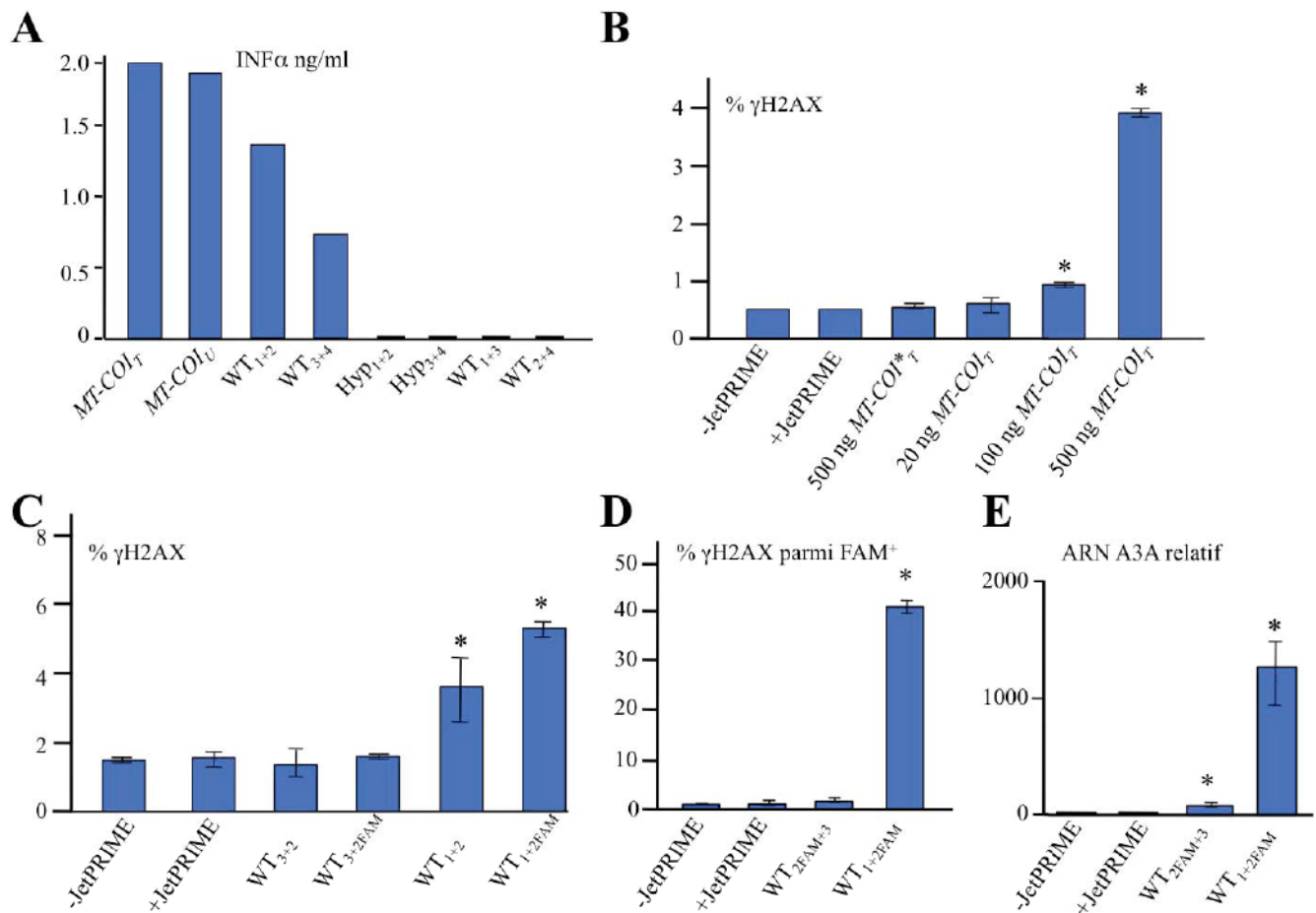


Figure 52 : L'ADN transfecté induit des cassures de l'ADN double brin.

A) Production d'interféron α des THP-1 transfectées avec 500ng de MT-COI_T ou MT-COI_U, 500ng de WT1+2, WT2+3, Hyp1+2, Hyp3+4, WT1+3 et WT2+4 (Tableau 1). B) Analyse par FACS des cellules positives pour γ H2AX transfectées avec l'ADN MT-COI. Les moyennes et les écarts types ont été déterminés à partir de deux transfections indépendantes. C) Analyse par FACS des cellules THP-1 positives pour γ H2AX transfectées avec WT3+2FAM, WT3+2FAM, W1+2FAM et W1+2FAM respectivement. Les moyennes et écarts types ont été déterminés à partir de deux transfections indépendantes. D) Analyse par FACS des cellules THP-1 positives pour γ H2AX parmi les cellules FAM-positives après transfection de WT3+2FAM et W1+2FAM respectivement. E) Profil transcriptionnel d'A3A dans les THP-1 transfectées avec WT1+2FAM (ADN double brin) ou WT3+2FAM (ADN simple brin) respectivement, en absence ou en présence de JetPRIME. * indique une différence statistiquement significative entre deux pourcentages observés ($p < 0,05$).

formes hypermutées de WT1+2 et WT3+4 respectivement (Tableau 4). Nous avons transfecté les cellules THP-1 avec des hétéroduplexes et nous avons mesuré la production d'IFN α (Figure 52A). WT1+2 et WT3+4 (ADN double brin), de façon similaire au fragment d'ADN *MT-COI*, induisent efficacement l'IFN α tandis que les couples non complémentaires, et donc simple brin WT1+3 et WT3+4 n'induisent pas de réponse IFN α (Figure 52A). L'absence d'induction d'IFN α par les couples complémentaires Hyp1+2 et Hyp3+4 peut s'expliquer par l'incapacité de ces couples à former des hétéroduplexes à 37°C. Ces résultats suggèrent fortement que l'ADN double brin est nécessaire à l'induction d'IFN. En effet, la transfection d'ADN *MT-COI* dans les THP-1 conduit à une augmentation significative des DSBs (Figure 52B), comme cela était attendu au vu de la surexpression d'A3A induite par l'IFN.

Afin de sélectionner spécifiquement les cellules transfectées, nous avons utilisé de l'ADN couplé au fluorphore FAM, W1+2FAM (ADN double brin) ou WT3+2FAM (ADN simple brin). 24h après transfection, on observe que l'ajout de FAM n'induit pas de DSB (WT3+2 vs WT3+2FAM et WT1+2 vs WT1+2FAM, Figure 52C). Néanmoins, la fluorescence FAM diminuant avec le temps, des cellules FAM+ ont été triées après 6 heures de transfection et ont été remises en culture 18h avant l'analyse des DSB par FACS. On observe que la transfection de la sonde double brin conduit à environ 40% de cellules THP-1 positives pour γ H2AX (Figure 52D), ce qui est corrélé à une surexpression d'A3A (Figure 52E). Ceci confirme donc l'impact de la transfection d'ADN simple brin sur l'expression d'A3A et la formation de DSB.

7. L'ADN double brin est transcrit par l'ARN polymérase III, ce qui déclenche la voie de signalisation RIG-I

Dans le but de déterminer quel senseur de l'ADN cytosolique déclenche la réponse interféron, nous avons transfecté l'ADN *MT-COI* dans la lignée THP-1. Après 48h de transfection, nous avons analysé par Western blot l'expression de protéines impliquées dans la détection d'ADN cytoplasmique. Nous pouvons observer une augmentation du niveau d'expression des formes phosphorylées de TBK1 (TBK1-P) et d'IRF3 (IRF3-P). Nous observons aussi une augmentation de RIG-I et MDA5 (Figure 53A) en présence d'ADN. En revanche, les niveaux de STING, IKK ϵ et MAVS restent inchangés. L'augmentation dose-

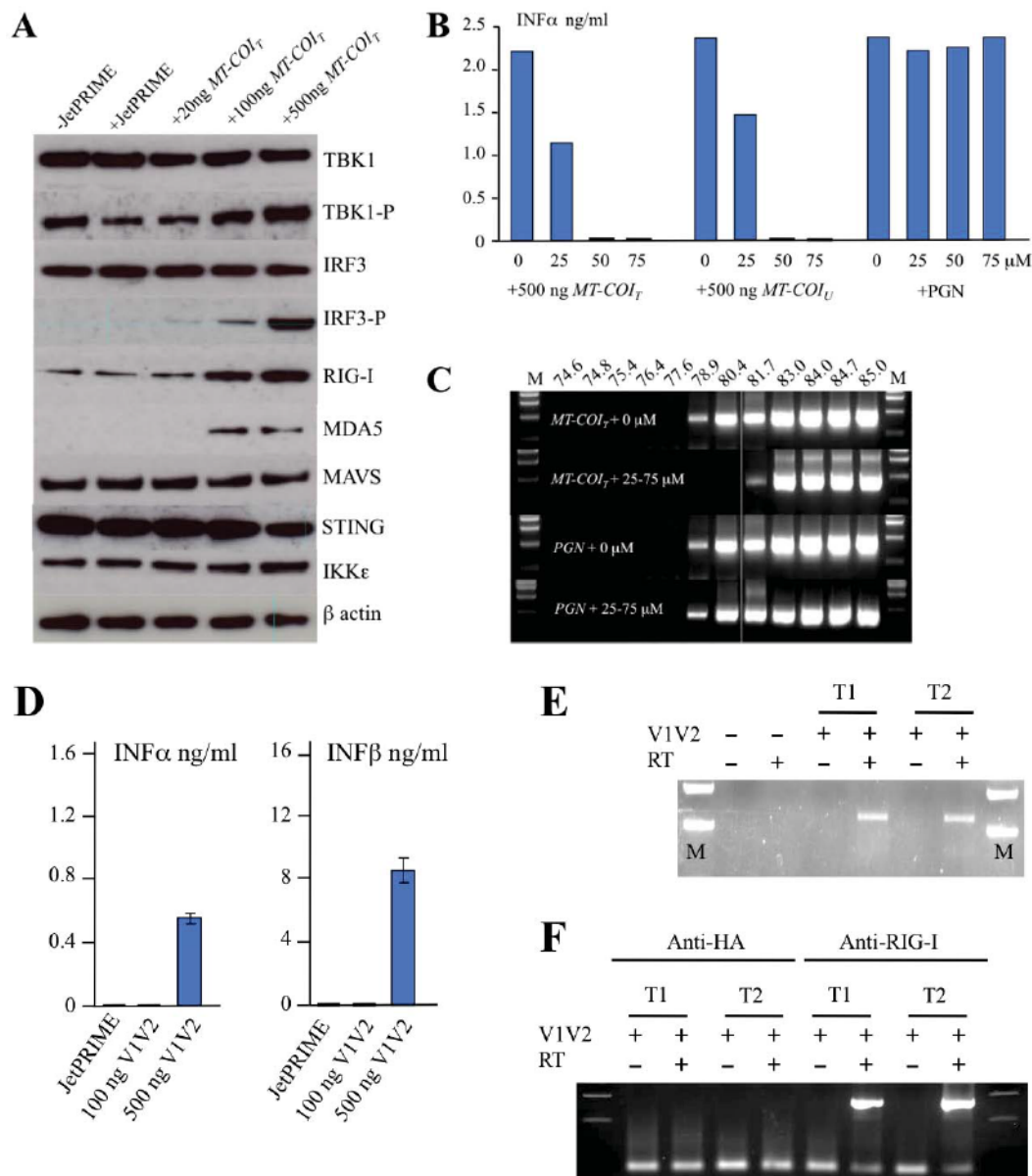


Figure 53: L'ADN double brin transfecté active la voie de signalisation RIG-I via l'ARN polymérase III.

A) Western blots sur TBK1, TBK1-P (phosphorylé), IRF3, IRF3-P (phosphorylé), RIG-I, MDA5, MAVS, STING, IKKε a 24 heures après transfection d'ADN dans les THP-1. B) Production d'interféron α par les THP-1 transfectées avec de l'ADN et traitées avec un inhibiteur d'ARN Polymérase III (ML-60218). Le Peptidoglycans (PGN) sert de contrôle. C) Editing par les A3 de l'ADN *MT-CYB* amplifié par 3D-PCR en absence ou en présence de 25, 50 ou 75 μM d'inhibiteur d'ARN polymérase III. La ligne blanche indique le seuil entre les amplicons édités et non édités. M : marqueur de poids moléculaire. D) Production d'interféron α et β par les THP-1 après transfection de 500ng d'ADN de V1V2 de VIH. Le JetPRIME sert ici de contrôle. E) Transcrits ARN de V1V2 issus de l'ADN transfecté dans les THP-1. L'ARN total a été extrait et le cDNA correspondant à V1V2 a été synthétisé en présence ou en absence de reverse transcriptase. T1 et T2 indiquent deux transfections indépendantes. F) Immunoprécipitation à l'aide d'un anticorps monoclonal anti-RIG-I ou d'un anticorps contrôle anti-HA. Les produits de RT-PCR de V1V2 ne sont obtenus que lorsque l'anticorps anti-RIG-I est employé. RT, T1 et T2 ont la même signification qu'en E.

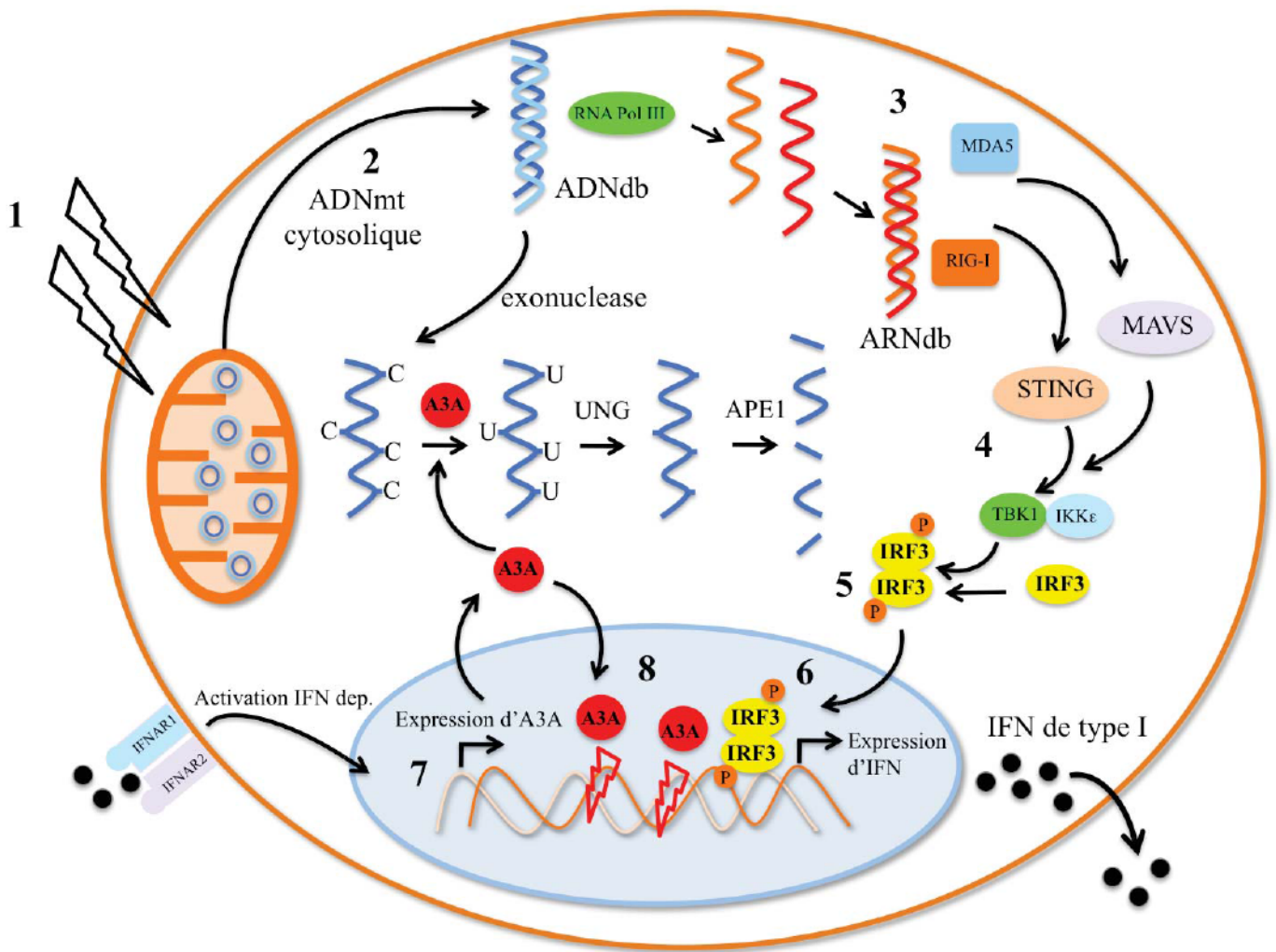


Figure 54 : Voie de signalisation détectant l'ADN cytosolique et déclenchant la réponse immunitaire.

Lors d'un stress cellulaire (1), l'ADNmt est relâché dans le cytoplasme (2) et est reconnu par l'ARN polymérase III. Il est alors transcrit en de nombreux ARN qui s'hybrident en duplexes d'ARN double brin (3). Ceci active la voie de signalisation RIG-I entraînant la phosphorylation d'IRF-3 par TBK1 (4-5) et la production d'interféron (6). Par l'intermédiaire d'IFNAR1/2, la production d'IFN induit la surexpression d'A3A (7), qui initie le catabolisme de l'ADN simple brin cytoplasmique. Etant donné qu'A3A entre aussi dans le noyau, elle peut induire des hypermutations de l'ADN nucléaire et la formation de DSB (8).

C. Conclusion

Les cytidines désaminases sont fortement exprimées dans la lignée hématopoïétique (Koning et al., 2009; Peng et al., 2007; Thielen et al., 2010). Elles peuvent induire, à l'exception d'A3B, des mutations de l'ADNcymt, ce qui est cohérent leur localisation cellulaire. Nous avons montré que la proportion de cellules dont l'ADNcymt est muté varie de 0,6 à 25% et peut augmenter lors d'un stress génotoxique, tandis que 1 à 4% des ADNcymt sont mutés dans ces cellules. Le réseau mitochondrial étant très variable entre les cellules en prolifération et en quiescence, le nombre de génomes mitochondriaux peut varier de 3 000 à 10 000 par cellules. Le nombre de fragments d'ADNcymt hyperédités par les A3 est donc de 30 à 400 par cellules (3 000-10 000 x 1-4%). De façon intéressante, la proportion de séquences *MT-COI* hypomutées est constamment de 50%. Etant donné que les enzymes A3A n'ont pas accès à la mitochondrie (Suspène, Aynaud, Guétard, et al., 2011) et n'éditent que l'ADN simple brin, ceci suggère que la moitié des génomes d'ADNmmt sont dans le cytoplasme sous forme fragmentée.

De plus, nous montrons ici que l'ADN transfecté peut être transcrit par l'ARN polymérase III conduisant à la formation d'ARN double brin. Ceux-ci sont détectés par RIG I, ce qui entraîne une réponse interféron et la surexpression d'A3A (Figure 54). Ce mécanisme est perdu lors de l'emploi d'un inhibiteur de l'ARN polymérase III, confirmant le rôle central de celle-ci. Nous avons pu observer que les uridines dans l'ADN ne constituent pas un signal de danger.

Les fragments d'ADN édités AU-riches sont thermodynamiquement moins stables : ils ont par conséquent moins de chance de se réhybrider après l'action des exonucléases. L'*editing* peut donc représenter un moyen pour la cellule d'éviter la formation de fragments d'ADN double brin dans le cytoplasme, perçus comme un signal de danger et constituant donc une source importante de stress.

Néanmoins, cette surexpression d'A3A n'est pas sans conséquence pour la cellule car ceci aboutit à l'*editing* de l'ADN chromosomique et à l'apparition de DSB. Ces deux phénomènes étant caractéristiques de certains cancer, on peut se demander quel est l'impact d'un tel mécanisme dans les situations d'inflammation chronique ou de maladies auto-immunes, dont certaines, telles que les interferonopathies et le lupus érythémateux, sont

dépendante de TBK1-P et d'IRF3-P, des éléments clés dans l'induction d'IFN, suggère que ceux-ci participent à un mécanisme aboutissant à la production d'IFN.

Il a été montré que l'ARN polymérase III est impliqué dans la transcription de l'ADN double brin cytosolique dsDNA (Ablasser et al., 2009; Y. H. Chiu et al., 2009). Nous nous sommes par conséquent intéressés au rôle de cette enzyme en transfectant la lignée THP-1 avec de l'ADN *MT-COI_T* ou *MT-COI_U* et en traitant les cellules avec du ML-60218, un inhibiteur de l'ARN polymérase III (L. Wu et al., 2003) (Figure 53B). Nous observons une corrélation inverse entre la production d'IFN α et la concentration d'inhibiteur employé. Afin d'exclure tout effet toxique de l'inhibiteur, nous avons traité les cellules avec 20 μ g/ml de peptidoglycanes (PGN) connus pour induire la production d'IFN α *via* le TLR2. Nous observons que, dans ce cas-là, l'inhibiteur n'a aucun impact sur la production d'IFN (Figure 53B). De plus, les niveaux d'expression de *RPL13A* mesurés par qRT-PCR quantitative ne sont pas affectés par l'inhibiteur, témoignant une nouvelle fois de la non-toxicité de la drogue aux concentrations utilisées. L'emploi de l'inhibiteur abolit l'*hyperediting* par les A3 de l'ADN *MT-CYB* provoqué par la transfection d'ADN *MT-COI_T* (Figure 53C). Encore une fois, l'emploi de PGN seul ou combiné au ML-60218 ne permet pas d'obtenir ce résultat (Figure 53C).

Si la détection de l'ADNcymt passe par la voie RIG-I après la transcription par l'ARN Polymérase III, il doit être possible de mettre en évidence de tels transcrits. Dans notre modèle, un fort bruit de fond généré par les ARN mitochondriaux endogènes est à craindre. Afin de pallier à ce phénomène, nous avons transfecté les THP-1 avec des produits de PCR correspondant à un segment de la séquence de l'enveloppe du VIH-1 (V1V2, 685 pb, 38% GC, (Suspène et al., 2013)). La transfection de 500ng d'ADN V1V2 induit une réponse IFN α/β comparable à celle obtenue avec l'ADN *MT-COIT* (Figure 53D vs Figure 49B et 49C). Nous avons extrait les ARN totaux à partir de deux transfections indépendantes, traité nos extraits à la DNase et réalisé une RT-PCR. Les ARN V1V2 intermédiaires ne sont obtenus que lorsqu'une étape de reverse transcription est réalisée (Figure 53E). Nous avons ensuite réalisé une expérience d'immunoprecipitation sur les lysats 8h post-transfection avec un anticorps monoclonal anti-RIG-I et un anticorps anti-HA qui sert ici de contrôle. Des produits de RT-PCR V1V2 sont obtenus seulement des immunoprecipitats anti RIG-I (Figure 53F).

caractérisées par un risque accru de cancers. Nous développerons cette question lors de la discussion générale.

En conclusion, nous avons montré que l'ADN cytosolique endogène et exogène constitue un signal de danger aboutissant à l'expression d'A3A pour cataboliser cet ADN, faisant donc d'A3A une enzyme anti-inflammatoire.

DISCUSSION

Lors de ces travaux, nous nous sommes intéressés à l'activité des A3 dans le contexte de deux stress cellulaires différents, un stress viro-induit : l'infection par le VHB, et un stress endogène : la présence d'ADN mitochondrial dans le cytoplasme. Ces deux stress se caractérisent par une inflammation importante aboutissant à l'expression des A3.

A. A3DE est-elle permise au VHB d'échapper à la restriction par A3G ?

Dans le contexte de l'infection par le VHB, nous avons observé, lors d'une étude précédente, la surexpression d'A3DE dans les tissus cirrhotiques (Vartanian et al., 2010). L'étude de cette A3, qui ne possède peu voire aucune activité cytidine désaminase a permis de mettre en évidence son rôle dans la modulation d'A3F et d'A3G. Ces deux A3 sont impliquées dans la restriction de nombreux virus tel que le VIH et le VHB. Dans ce dernier cas, il a été montré qu'A3G est le principal facteur de restriction du VHB. Bien que des génomes hypermutés soient retrouvés chez les patients, ce virus se réplique dans le foie et semble échapper à une restriction totale par A3G. Il est donc légitime de penser qu'une des stratégies mises en place par le VHB pour échapper à la restriction d'A3G est de se répliquer à faible bruit afin d'éviter de déclencher une réponse immunitaire trop forte et d'entraîner une surexpression d'A3G suite à la production d'IFN γ . Cette hypothèse est cependant contredite par le fait que les transcriptomes de foie sain ou de lignées issues de cellules hépatiques montrent que le niveau basal d'A3G est important.

A l'heure actuelle il semble également que le VHB ne possède pas de protéine de type Vif permettant de moduler les protéines A3F et A3G. On peut donc se demander comment un virus comme le VHB peut se répliquer dans des cellules exprimant les A3F et A3G sans pouvoir moduler ces enzymes au moyen d'une protéine virale de type Vif.

La mise en évidence du rôle modulateur d'A3DE permet en partie de répondre à cette problématique. En effet, A3DE est aussi exprimée dans les cellules hépatiques et dans les lignées dérivées d'hépatocytes. De façon intéressante, on observe que le niveau basal d'A3DE est similaire à celui d'A3F et d'A3G. De cette façon, A3DE inhibe l'activité d'A3F et d'A3G à leur niveau basal. Cette inhibition étant dose dépendante, on peut envisager qu'une forte inflammation pourrait entraîner une forte surexpression d'A3F et d'A3G et permettre à celles-ci d'outrepasser l'inhibition par A3DE. Les infections par le VHB sont traitées à l'aide d'IFN α pégylé, la pégylation permettant d'augmenter la demi-vie de l'IFN. Ceci va contribuer à induire une expression d'A3G plus forte : on peut donc supposer qu'A3G contribue de cette

façon à l'élimination du virus. Aux vues des résultats obtenus, on peut supposer que cette stimulation est nécessaire malgré les niveaux importants d'A3G dans les cellules hépatiques du fait de l'inhibition de cette dernière par A3DE.

Ceci implique néanmoins qu'A3DE, appartenant à une famille de facteurs de restrictions viraux, a dans ce contexte un rôle proviral. De façon intéressante, il a été observé que d'autres A3 peuvent avoir un rôle proviral dans un contexte particulier. C'est le cas par exemple d'A3A qui améliore la réplication du virus West Nile et du virus de l'encéphalite équine vénézuélienne (Schoggins et al., 2011). C'est aussi le cas de la protéine ADAR1-L qui restreint le virus de la rougeole mais qui est nécessaire à la réplication du virus de l'hépatite delta (Polson, Bass, & Casey, 1996). On peut supposer qu'A3DE est nécessaire à la réplication du VHB. Le VIH possède la protéine Vif qui permet de contrecarrer la restriction par les A3. Or, aucune protéine similaire à Vif de VIH-1 n'a été mise en évidence pour le VHB. On peut supposer que le VHB, à défaut de posséder une protéine anti A3G, induit la surexpression d'A3DE afin d'inhiber sa restriction par A3F et A3G.

Même si actuellement nous n'avons pu mettre en évidence une surexpression d'A3DE lors de la transfection du plasmide pCayw dans la lignée Huh7. il est tout à fait envisageable que le virus induise la surexpression d'A3DE de façon indirecte et qu'il soit nécessaire d'étudier la surexpression d'A3DE dans un contexte plus physiologique afin de décrypter par quel mécanisme l'infection par le VHB peut mener à la surexpression de cet inhibiteur d'A3F et d'A3G

Par ailleurs, il a été montré qu'A3G induit aussi des hypomutations sur le génome du VHB, notamment au niveau des gènes pré-core et AgHBe. Ce phénomène peut contribuer à l'échappement du virus à la réponse immunitaire (Vartanian et al., 2010). En ce sens, A3DE peut contribuer à l'augmentation du *fitness* du virus : en modulant le niveau d'expression d'A3G, A3DE diminue le nombre de mutations sur le génome du VHB et favorise les phénomènes d'hypomutations. Un mécanisme similaire a déjà été mis en évidence pour le VIH-1 qui grâce à des protéines Vif plus ou moins fonctionnelles module le niveau d'expression et l'activité d'A3G à fin d'augmenter son taux de mutations et donc sa variabilité.

Enfin, il serait intéressant d'étudier le rôle d'A3DE dans le contexte d'autres infections virales impliquant une restriction par A3F et d'A3G. Dans le contexte de l'infection par le VIH par exemple, on peut imaginer que l'inhibition d'A3F et d'A3G facilite la réplication du virus et le rôle de Vif. Il serait intéressant aussi d'étudier l'infection par le VIH dans le contexte d'une active d'A3DE : A3DEC(320)Y. Cette forme n'est présente que chez 0,1% de la population mondiale, sans biais ethnique ou géographique. Il est donc peu probable que cet allèle soit lié à une évolution différente de la maladie. Néanmoins, il a été montré que A3DEC(320)Y induit des hypermutations sur le génome du VIH-1. Il est donc probable que malgré l'action de la protéine Vif, le génome viral soit plus édité chez les personnes possédant la forme active d'A3DE que chez les personnes possédant la forme non active. L'étude dans le modèle simien pourrait constituer un modèle intéressant étant donné qu'à l'exception du gorille, tous les singes possèdent une forme active d'A3DE.

Cette faible fréquence de l'allèle C320Y chez l'Homme peut aussi être le résultat d'erreurs de séquençages. En effet ces techniques de séquençage ne sont pas adaptées à un cluster dont les gènes le composant sont si proche phylogénétiquement favorisent ainsi les erreurs par hybridation des amorces sur des séquences d'autres gènes proches. Ainsi, il est probable que les séquences d'A3DE étant caractérisés C320Y soient en réalité le produit de séquençage d'A3F, qui est très proche d'A3DE et qui possède un Y en position 320. Il se pourrait donc que la forme active n'existe pas chez l'Homme et ne soit qu'une construction artificielle. Une étude plus fine est donc nécessaire afin de préciser la réelle fréquence de la forme active chez l'Homme.

Ces données apportent des éléments nouveaux sur la compréhension de l'échappement du VHB à la restriction par les A3. Il pourrait être intéressant d'envisager des méthodes thérapeutiques visant à bloquer A3DE et ainsi permettre à A3F et A3G de restreindre le virus du VHB.

B. Quel est le rôle physiologique d'A3DE ?

L'étude des homologues d'A3DE chez les primates s'avère intéressante. En effet l'allèle actif, A3DEC(320)Y, est retrouvé chez les singes du nouveau et de l'ancien monde, à l'exception du gorille et de l'Homme, chez qui l'allèle non actif A3DEC(320) est retrouvé. A3DE est particulièrement active chez le chimpanzé, sur le VIH notamment. Ceci signifie que

la perte d'activité serait spécifique à l'Homme et au gorille. Il est néanmoins nécessaire d'étudier les A3DE chez plusieurs espèces de singe afin de déterminer si celle-ci est bien active chez ces espèces. A3H est une autre A3 à avoir perdu son activité récemment puisqu'il s'agit d'une A3 très active chez les singes mais dont il existe des allèles codant pour des formes peu stable et donc peu actives chez l'Homme (OhAinle et al., 2008). A3DE et A3H sont donc deux A3 ayant perdu leur activité spécifiquement chez l'Homme.

Il est possible que cette perte d'activité soit le résultat d'une absence de pression de sélection sur l'A3DE humaine. On peut alors envisager que les chimpanzés et les autres singes possédant un A3DE fonctionnelle aient besoin de cette cytidine désaminase active comme moyen de défense contre un virus ou alors dans une fonction physiologique particulière. Etant donné que d'autres A3 ont conservé une forte activité, il est possible qu'A3DE, aussi bien qu'A3H, étaient initialement impliqués dans la restriction d'un virus éteint chez l'Homme mais toujours actif chez le singe. Des études sur les A3DE de singe permettraient de mieux comprendre l'évolution d'A3DE et plus globalement du cluster APOBEC3.

L'hypothèse qu'il existe un bénéfice à posséder une forme d'A3DE non active n'est néanmoins pas à exclure. Nous avons montré que la forme A3DEC(320)Y est active sur le VHB et qu'elle interagit aussi avec A3F et A3G. L'interaction n'est donc pas le produit de la perte d'activité mais une propriété ancestrale d'A3DE, conférant à celle-ci une nouvelle fonction lorsqu'elle est inactive. Il serait intéressant d'étudier la fonction d'A3DE chez les singes possédant l'allèle actif et de vérifier dans quels contextes celle-ci est un avantage ou est délétère pour l'organisme.

Cette sélection récente d'une forme non active laisse penser qu'une pression de sélection particulière ou un « bottleneck » évolutif est à l'origine de cette perte d'activité (Gagneux et al., 1999). Ceci est compatible avec une fonction d'A3DEC(320)Y conférant un avantage sélectif dans un environnement particulier mais étant délétère dans d'autres circonstances, aboutissant à une contre sélection de la forme active.

Si la forme inactive confère un avantage, il est nécessaire de déterminer le rôle de cette inhibition. Il a été montré qu'A3F et A3G sont capables d'induire des mutations sur l'ADN cytoplasmique (plasmidique ou ADN mitochondrial cytoplasmique). A3DE pourrait avoir

pour fonction d'éviter qu'A3F et A3G n'induisent des mutations sur ces cibles en dehors d'un stress cellulaire. On peut supposer que ceci permet de protéger l'ADN chromosomique lors de la mitose par exemple ou bien d'éviter l'editing des ADN extrachromosomaux circulaires, bien que le rôle de ces derniers soit encore peu connu. L'interaction d'A3DE avec A3F et A3G pourrait aussi être liée à la fonction de ces dernières dans les Processing bodies (P-bodies). Ces structures cytoplasmiques sont le siège de la dégradation et du *decapping* des ARNm notamment. Il a été montré qu'A3DE, A3F et A3G sont retrouvées dans ces structures. Il est possible qu'A3DE ait pour fonction d'inhiber l'activité d'A3F et d'A3G en dehors de ces structures. Il en est de même pour les complexes de haut poids moléculaire composés d'ARNm cellulaires auxquels A3G est associé. De façon intéressante, l'étude des transcriptomes de différentes lignées cellulaires et tissus montre qu'A3DE est quasi systématiquement exprimée avec A3F et A3G, bien que leurs promoteurs diffèrent. Ceci renforce l'idée qu'A3DE a un rôle fondamental dans la régulation de l'activité basale d'A3F et d'A3G.

Enfin, de récents travaux montrent que A3A peut être impliquée dans des mutations de sur des ARNm, notamment sur le transcrit *WT1* (Niavarani et al., 2015). Ce gène code pour une protéine régulatrice pouvant à la fois être oncogène ou suppresseur de tumeur selon l'isoforme considéré. Ainsi, cette étude établit la preuve de concept que les A3 peuvent potentiellement éditer des acides ribonucléiques. Il est possible que la forme active d'A3DE, ainsi que d'autres A3, soient impliquée dans la désamination d'ARNm. D'autant plus qu'A3F et A3G peuvent se fixer à l'ARN. Etant donné la proximité entre A3DE et A3F, il est probable que la forme active d'A3DE puisse aussi se fixer à l'ARN. Ainsi, on peut imaginer qu'elles possèdent une fonction de désamination sur des ARNm de façon régulées, de la même manière qu'A1 sur l'ARNm de l'ApoB. De la même façon, ceci laisse supposer que les A3 peuvent aussi muter le génome de virus à ARN bien que le seul l'*editing* par ADAR-1 a été mis en évidence pour ces virus aujourd'hui.

C. Un rôle physiologique pour A3A ?

Si l'activité antivirale de certaines A3 n'est plus à démontrer il semble cependant que cette activité ne soit pas le rôle principal de ces enzymes et que certaines possèdent un véritable rôle physiologique. Récemment nous avons pu mettre en évidence qu'A3A joue un

rôle clé dans le catabolisme de l'ADN en induisant sa dégradation par la machinerie cellulaire sans pour autant établir le mécanisme exact.

L'ADNmt se retrouvant dans le cytoplasme est une source de stress pour la cellule. Il est donc nécessaire que celui-ci soit dégradé. Le mécanisme de catabolisme de l'ADN cytosolique que nous proposons est le suivant : l'ADN double brin est d'abord dégradé sous forme de fragments simple brin par TREX1, ou d'autres endo- et exo- nucléases. Le rôle d'A3A serait alors d'induire des désaminations sur les fragments simple brin qui seraient ensuite fragmentés après l'action de l'UNG et d'APE1. A3A empêcherait ainsi l'hybridation des fragments générés par TREX1, qui pourraient dans le cas contraire à nouveau stresser la cellule par la voie de l'ARN polymérase III et de RIG-I. Cette fonction d'A3A est fondamentale puisque de l'ADN édité est retrouvé dans de nombreux types cellulaires, mais aussi dans la circulation chez des patients sains et atteints de sepsis. De plus, A3A est conservée dans l'évolution depuis 150 millions d'années (Caval, Suspene, Vartanian, et al., 2014). Ces éléments tendent à montrer qu'A3A est impliquée dans un processus biologique très différent de ceux qui lui étaient attribués jusqu'ici.

Afin de compléter ces observations, il est néanmoins nécessaire d'étudier le catabolisme de l'ADNcymt dans d'autres lignées cellulaires. En effet, A3A est inductible dans la lignée myéloïde et dans les kératinocytes principalement (Refsland et al., 2010). Quand est-il de la dégradation de l'ADNcymt dans les lignées dépourvues d'A3A ? Il est envisageable que d'autres A3 soient chargées de la dégradation de l'ADNmtcyt, tel que A3F et A3G. La recherche de mutations imputables à A3G dans ces lignées serait aisée du fait du contexte d'*editing* particulier d'A3G. D'autres A3 tel qu'A3C, dont le rôle n'est pas encore bien connu, pourraient être impliquées dans ce processus. Le fait qu'A3A soit sollicitée témoigne néanmoins d'un contexte particulier dans la lignée myéloïde, nécessitant l'expression rapide d'une A3 à forte activité tel qu'A3A. Ceci peut être lié aux conditions de stress auxquels sont soumises ces cellules notamment lié à leur rôle dans l'inflammation.

De plus, il est maintenant établi que des séquences d'ADN mitochondrial peuvent entrer dans le noyau et s'insérer dans l'ADN génomique. Ces séquences sont appelées NUMT pour *Nuclear Mitochondrial DNA*. On peut supposer que l'importance de la dégradation de l'ADN mitochondrial soit liée à ce phénomène. En effet, l'insertion de séquences pourrait avoir un effet délétère pour le génome. Il est donc nécessaire de réguler ce mécanisme en

dégradant efficacement les fragments d'ADN cytoplasmique d'origine mitochondrial. Dans ce contexte, il apparaît qu'A3A contribue à maintenir la stabilité génomique en participant à la dégradation de l'ADNcymt, d'autant plus qu'une étude récente montre que dans certains cancers, les insertions d'ADNmt dans l'ADN génomique sont plus fréquentes dans les cellules tumorales que dans les cellules saines (Ju et al., 2015). Il serait intéressant de caractériser la nature des mutations présentes sur ces séquences. La découverte d'hypomutations ou d'hypermutations dans le contexte 5'TpC apporterait un argument supplémentaire quant au rôle d'A3A dans la dégradation de l'ADNcymt.

Ces observations ouvrent aussi de nouvelles voies d'étude sur le rôle d'A3A dans la restriction bactérienne. En effet, étant donné qu'A3A peut induire des mutations sur l'ADN cytoplasmique dans les THP-1, lignée dérivée de monocytes, il serait d'intérêt d'étudier le rôle de cette cytidine désaminase dans le contexte d'infection par des bactéries à répllication intracellulaire tel que *Chlamydia* ou *Rickettsia*. On peut supposer qu'A3A va participer à l'élimination de ces bactéries. La recherche de génomes hypermutés dans des macrophages infectés en culture ou chez des patients permettrait de déterminer le rôle d'A3A dans ce contexte. Il en est de même pour les parasites intracellulaire infectant les macrophages, tel que *Leishmania*

D.A3A, un facteur aggravant dans les maladies liées au catabolisme de l'ADN mitochondrial ?

Nous avons vu que l'activité catabolique d'A3A peut s'accompagner d'une génotoxicité pour la cellule en cas de surexpression massive de cette enzyme. Il existe donc une étroite frontière pour cette enzyme qui dans son rôle physiologique remplit une fonction anti inflammatoire mais qui en cas de surexpression peut entraîner des dommages sur le génome.

A la lumière de ces résultats, il serait intéressant d'étudier l'editing de l'ADNmt et le rôle d'A3A chez les patients atteints du syndrome d'Aicardi-Goutières. Ce syndrome peut être causé par une déficience pour l'enzyme TREX1. Parmi les symptômes associés à ce syndrome, on retrouve une accumulation d'ADN cytoplasmique et une forte production d'IFN-I et de leucocytes, notamment dans le liquide céphalo rachidien. Dans ce contexte, on

peut imaginer que l'ADN cytoplasmique, n'étant pas dégradé par TREX1, stimule en continue la voie RIG-I, aboutissant à la forte production d'IFN-I observée et donc à une forte surexpression d'A3A. Dans ce sens, il serait intéressant de rechercher des mutations sur l'ADNmt de ces patients et d'en évaluer la fréquence. Celle-ci devrait être largement supérieure à celle observée chez des patients sains. Le taux de mutations sur l'ADN nucléaire et les DSB devraient aussi être très important chez ces patients. Le gène *TREX1* est aussi muté dans certains lupus érythémateux. Cette maladie se caractérise aussi par une forte inflammation et l'accumulation d'ADN cytoplasmique. Le lupus favorisant l'apparition de cancers, notamment de lymphomes, on peut se demander si dans ce cas, la forte expression d'A3A, résultant de l'accumulation d'ADN cytoplasmique et de la production d'IFN-I, ne contribue pas à la tumorigénèse par l'induction de mutations nucléaires et de DSB. D'une manière plus générale, cette problématique est valable pour toutes les maladies auto immunes s'accompagnant d'une inflammation importante.

Nous avons évoqué précédemment le cas des NUMTs. Il a été montré que ceux-ci s'intègrent surtout au niveau des sites de cassures de l'ADN double brin (Hazkani-Covo, Zeller, & Martin, 2010). Dans le cas d'une lésion sur le gène *TREX1*, la persistance d'ADNmt dans le cytoplasme pourrait éventuellement favoriser l'intégration de ceux-ci dans le génome nucléaire. Or, cette accumulation d'ADNmt entraîne une forte inflammation cellulaire par le biais de la production de l'IFN et aboutissant à la surexpression d'A3A. Nous avons montré que dans ces conditions, A3A peut induire des DSB. Ainsi, l'accumulation d'ADNcytmt et la formation de DSB pourraient favoriser le phénomène de NUMTs. Dans cette optique, il pourrait être judicieux d'étudier la proportion de NUMTs chez des patients atteints de lésions sur le gène *TREX1* et de la comparer à celle de patients sains, notamment dans les cellules exprimant fortement A3A. Dans ces conditions, il est probable qu'A3A participe à une certaine instabilité génétique. Ceci suppose néanmoins que l'activité d'A3A ne déclenche pas d'apoptose.

Par ailleurs, l'étude de la délétion $\Delta A3B$ est aussi un sujet d'étude pertinent. De façon intéressante, de nombreuses études ont montré que l'incidence et la mortalité du lupus érythémateux systématique est plus important chez les populations chinoises et aborigènes (Anstey, Bastian, Dunckley, & Currie, 1993; Segasothy & Phillips, 2001). Ces mêmes populations possèdent la fréquence la plus élevée de délétion $\Delta A3B$, dépassant 90% chez les aborigènes. Ces résultats supportent l'hypothèse qu'A3A pourrait avoir un rôle très délétère

chez les patients lupiques : la forme d'A3A codée par les individus $\Delta A3B$ étant plus active, il est probable que la surexpression d'A3A chez les patients lupiques entraîne une plus forte mortalité chez les patients lupiques $\Delta A3B$. Une étude épidémiologique à ce niveau permettrait de mieux comprendre la façon dont le lupus peut avoir des effets différents selon les patients. Elle permettrait aussi un dépistage plus efficace en considérant cette délétion comme un facteur de risque, et donc une prise en charge plus soutenue de ces individus.

Enfin, il serait intéressant d'étudier l'expression de TRIB3, un régulateur de l'activité d'A3A lorsque la cellule est stressée par l'ADNcymt. En effet, TRIB3 induit la dégradation d'A3A ce qui se traduit par une diminution de l'*editing* nucléaire. TRIB3 étant une protéine strictement nucléaire, on peut supposer que celle-ci permet d'éviter l'*editing* du génome cellulaire lorsque celle-ci catabolise l'ADNcymt par le biais d'A3A. L'activité catabolique d'A3A n'est donc pas en contradiction avec la survie cellulaire mais nécessite une régulation fine et complexe. Ceci expliquerait en partie pourquoi un tel processus biologique ait été sélectionné par l'évolution : garder un puissant mutateur comme A3A permet d'éliminer efficacement une source de stress cellulaire tout en prenant soin, par l'intermédiaire de TRIB3, de limiter l'activité d'A3A dans le noyau.

E. Quelle est la contribution respective d'A3A et d'A3B dans la carcinogénèse ?

Récemment, plusieurs groupes ont proposé qu'A3B soit la principale cytidine désaminase impliquée dans la carcinogénèse (Burns, Lackey, et al., 2013; Burns, Temiz, et al., 2013; Shinohara et al., 2012). Ces études sont principalement basées sur l'analyse des mutations présentes dans le génome de cancers (seins, ...) et sur des analyses transcriptionnelles de ces cancers. En effet A3B est la seule cytidine désaminase à être fortement exprimée dans les cellules cancéreuses et dans les lignées cellulaires cancéreuses contrairement à A3A, même si aucune de ces études n'a permis de démontrer qu'A3B induit des mutations dans l'ADN génomique. En effet les mutations du génome attribuées à A3B dans ces études sont en réalité des erreurs techniques liées à la 3D-PCR.

Nous avons été les premiers à mettre en évidence qu'A3B peut en effet induire des mutations de l'ADN nucléaire sans pour autant être capable d'induire des cassures de l'ADN chromosomique (Caval, Suspene, Shapira, et al., 2014). Il est donc probable qu'A3B comme A3A puissent contribuer à la carcinogénèse. Cependant plusieurs éléments nous conduisent à penser que contrairement à A3A qui joue un rôle majeur dans la carcinogénèse, celui d'A3B est à relativiser.

Si l'on compare les profils d'expression d'A3A et d'A3B, ceux-ci sont totalement opposés. En effet, A3A est très faiblement voire pas exprimé dans la majorité des cellules en dehors de celles issues de la lignée myéloïde, contrairement à A3B qui est exprimé dans la majorité des cellules à un niveau relativement important. A3A a une localisation nucléocytoplasmique tandis qu'A3B est uniquement nucléaire (Refsland et al., 2010). A3A est une protéine dont l'expression est fortement induite par l'IFN (la surexpression peut augmenter de 1000 fois) contrairement à A3B dont l'expression n'est que très faiblement affectée par l'IFN.

Dans le cas d'A3B, il nous apparaît difficilement conciliable que la majorité des cellules expriment de façon constitutive et à un taux relativement important une protéine nucléaire qui se comporte comme un mutateur de l'ADN, ou alors il est nécessaire que cette protéine soit extrêmement régulée dans le noyau des cellules, notamment au moment de la division cellulaire. Il est à noter que pour le moment, aucune protéine régulatrice d'A3B n'a été identifiée ce qui semble contredire cette hypothèse.

A3A en revanche est très faiblement exprimée voire pas du tout dans la majorité des cellules mais son expression est extrêmement inductible par l'IFN. Il est donc légitime de penser qu'il existe des « vagues d'expression » d'A3A dans les cellules suite à l'induction par différents types de stress (infection virales, stress physique ou chimique), ce qui pourrait conduire à des « vagues de mutations » dans l'ADN génomique. Même si la majorité des cellules subissant ces « vagues mutationnelles » entrent en apoptose ou que la majorité des mutations soient corrigées par les systèmes de correction des erreurs, on peut imaginer qu'un certain nombre de mutations puissent subsister et entraîner certaines cellules vers la carcinogénèse.

Plusieurs données récemment publiées viennent renforcer notre modèle. Tout d'abord il est à noter que le gène codant A3B est délété chez 30 à 40% de la population d'Asie du Sud-Est (Kidd et al., 2007). Cette délétion de 29, 5 kb d'A3B est actuellement considérée comme un facteur de risque dans le développement des cancers du sein, des ovaires ou du foie, principalement pour les individus homozygotes pour la délétion d'A3B (T. Zhang et al., 2012). Pour un même type de cancer il est à noter que la fréquence des mutations attribuables aux A3 détectées dans les génomes issus de patients homozygotes pour la délétion d'A3B est supérieure à celle détectée dans les génomes de patients ne portant pas cette délétion.

Quelle est donc la source des mutations attribuables aux A3 chez les personnes délétées d'A3B ? Nous avons récemment pu apporter un premier élément de réponse en démontrant comment la délétion d'A3B contribue à l'augmentation de la fréquence des mutations attribuables aux A3 détectées chez les patients délétés d'A3B.

En effet, nous avons démontré que la délétion de 29, 5 kb d'A3B conduit à la production d'un ARNm d'A3A possédant la partie 3'UTR d'A3B ce qui conduit à la production d'un ARNm chimérique d'A3A beaucoup plus stable que la forme normale. La conséquence directe est la production d'une quantité plus importante d'A3A et donc une génotoxicité potentiellement accrue.

Nous avons validé expérimentalement ce modèle dans la lignée cellulaire SKBR3. Cette lignée dérivée de cancers du sein porte la délétion d'A3B et exprime donc un ARNm chimérique d'A3. L'induction de l'expression d'A3A par la PMA dans cette lignée aboutit à la formation de DSB. Cet exemple montre que des dommages de l'ADN sont possibles en absence d'A3B et illustre le risque lié à une activité d'A3A trop importante (Caval, Suspene, Shapira, et al., 2014).

Très récemment, une étude portant sur l'expression d'A3B dans des cancers du sein de haut et de bas grades a montré qu'A3B est bien surexprimée dans les tumeurs les plus agressives. Cependant, les auteurs ont montré que cette surexpression est en réalité une conséquence de la prolifération des cellules cancéreuses, et ont ainsi suggéré qu'A3B est surexprimée à certaines étapes du cycle cellulaire. A3B constitue en réalité un marqueur de la prolifération cellulaire et donc de l'agressivité de la tumeur mais n'est pas à l'origine de celle-

ci. Les auteurs de cette étude considèrent ainsi qu'il est peu probable qu'A3B soit à l'origine des mutations aboutissant aux cancers (Cescon, Haibe-Kains, & Mak, 2015).

Un des principaux arguments avancés par les détracteurs d'A3A dans la carcinogénèse est son faible niveau d'expression dans les cellules. L'absence d'expression d'A3A au stade de la tumeur peut être le résultat d'une contre sélection, on peut en effet supposer qu'A3A induit des mutations de type « *hit and run* » mais que l'expression de celle-ci soit « éteinte » au stade tumoral afin d'éviter l'induction de l'apoptose du fait de sa génotoxicité. Ainsi, les cellules tumorales exprimant peu ou pas d'A3A seraient sélectionnées au cours du temps. Le facteur temporel est donc fondamental dans l'étude de la carcinogénèse ; les mutations observées peuvent avoir une origine antérieure à la tumeur (pré tumorale) et le transcriptome tumoral n'équivaut pas au transcriptome pré tumoral. La carcinogénèse est un processus long, souvent engagé 10 à 20 ans avant l'apparition de la tumeur. Il en résulte que ce qui est observé au moment de l'expérience ne reflète pas tout le processus de développement du cancer. Il serait nécessaire de pouvoir suivre l'évolution du niveau d'expression des A3 tout au long du développement des cancers, ce qui malheureusement est techniquement impossible.

Si A3B ne contribue pas où que faiblement à la carcinogénèse son rôle physiologique n'a pas encore été clairement établi même si cette protéine semble être un marqueur de la prolifération cellulaire. Son absence chez 30 à 40% de la population d'Asie du Sud-Est suggère que cette protéine n'est pas nécessaire pour la survie ou la reproduction de l'espèce. L'origine de la délétion d'A3B dans ces populations n'a pour le moment pas d'explication scientifique mais semble conférer un avantage sélectif malgré le risque plus élevé de développer un cancer.

BIBLIOGRAPHIE

-A-

- Abe, H., Ochi, H., Maekawa, T., Hatakeyama, T., Tsuge, M., Kitamura, S., . . . Chayama, K. (2009). Effects of structural variations of APOBEC3A and APOBEC3B genes in chronic hepatitis B virus infection. *Hepatol Res*, 39, 1159-1168. doi: HEP566 [pii] 10.1111/j.1872-034X.2009.00566.x
- Ablasser, A., Bauernfeind, F., Hartmann, G., Latz, E., Fitzgerald, K. A., & Hornung, V. (2009). RIG-I-dependent sensing of poly(dA:dT) through the induction of an RNA polymerase III-transcribed RNA intermediate. *Nat Immunol*, 10(10), 1065-1072. doi: 10.1038/ni.1779
- Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., . . . Stratton, M. R. (2013). Signatures of mutational processes in human cancer. *Nature*, 500(7463), 415-421. doi: 10.1038/nature12477
- Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J., & Stratton, M. R. (2013). Deciphering signatures of mutational processes operative in human cancer. *Cell Rep*, 3(1), 246-259. doi: 10.1016/j.celrep.2012.12.008
- An, P., Johnson, R., Phair, J., Kirk, G. D., Yu, X. F., Donfield, S., . . . Winkler, C. A. (2009). APOBEC3B deletion and risk of HIV-1 acquisition. *J Infect Dis*, 200(7), 1054-1058. doi: 10.1086/605644
- Anstey, N. M., Bastian, I., Dunckley, H., & Currie, B. J. (1993). Systemic lupus erythematosus in Australian aborigines: high prevalence, morbidity and mortality. *Aust N Z J Med*, 23(6), 646-651.
- Argyris, E. G., Acheampong, E., Wang, F., Huang, J., Chen, K., Mukhtar, M., & Zhang, H. (2007). The interferon-induced expression of APOBEC3G in human blood-brain barrier exerts a potent intrinsic immunity to block HIV-1 entry to central nervous system. *Virology*, 367(2), 440-451. doi: S0042-6822(07)00413-8 [pii] 10.1016/j.virol.2007.06.010
- Armitage, A. E., Katzourakis, A., de Oliveira, T., Welch, J. J., Belshaw, R., Bishop, K. N., . . . Iversen, A. K. (2008). Conserved footprints of APOBEC3G on Hypermutated human immunodeficiency virus type 1 and human endogenous retrovirus HERV-K(HML2) sequences. *J Virol*, 82(17), 8743-8761. doi: JVI.00584-08 [pii] 10.1128/JVI.00584-08
- Ashur-Fabian, O., Har-Zahav, A., Shaish, A., Wiener Amram, H., Margalit, O., Weizer-Stern, O., . . . Rechavi, G. (2010). apoB and apobec1, two genes key to lipid metabolism, are transcriptionally regulated by p53. *Cell Cycle*, 9(18), 3761-3770.
- Atianand, M. K., & Fitzgerald, K. A. (2013). Molecular basis of DNA recognition in the immune system. *J Immunol*, 190(5), 1911-1918. doi: 10.4049/jimmunol.1203162
- Aynaud, M. M., Suspène, R., Vidalain, P. O., Mussil, B., Guétard, D., Tangy, F., . . . Vartanian, J. P. (2012). Human Tribbles 3 protects nuclear DNA from cytidine deamination by APOBEC3A. *J Biol Chem*, 287(46), 39182-39192. doi: 10.1074/jbc.M112.372722

-B-

- Baird, W. M., Hooven, L. A., & Mahadevan, B. (2005). Carcinogenic polycyclic aromatic hydrocarbon-DNA adducts and mechanism of action. *Environ Mol Mutagen*, 45(2-3), 106-114. doi: 10.1002/em.20095
- Baker, N. A., Sept, D., Joseph, S., Holst, M. J., & McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A*, 98(18), 10037-10041. doi: 10.1073/pnas.181342398

- Balkwill, F., & Mantovani, A. (2001). Inflammation and cancer: back to Virchow? *Lancet*, 357(9255), 539-545. doi: 10.1016/S0140-6736(00)04046-0
- Baumert, T. F., Rosler, C., Malim, M. H., & von Weizsacker, F. (2007). Hepatitis B virus DNA is subject to extensive editing by the human deaminase APOBEC3C. *Hepatology*, 46(3), 682-689.
- Beale, R. C., Petersen-Mahrt, S. K., Watt, I. N., Harris, R. S., Rada, C., & Neuberger, M. S. (2004). Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo. *J Mol Biol*, 337(3), 585-596.
- Bennett, R. P., Salter, J. D., Liu, X., Wedekind, J. E., & Smith, H. C. (2008). APOBEC3G subunits self-associate via the C-terminal deaminase domain. *J Biol Chem*, 283(48), 33329-33336. doi: 10.1074/jbc.M803726200
- Berger, G., Durand, S., Fargier, G., Nguyen, X. N., Cordeil, S., Bouaziz, S., . . . Cimorelli, A. (2011). APOBEC3A is a specific inhibitor of the early phases of HIV-1 infection in myeloid cells. *PLoS Pathog*, 7(9), e1002221. doi: 10.1371/journal.ppat.1002221
- Bishop, K. N., Holmes, R. K., Sheehy, A. M., Davidson, N. O., Cho, S. J., & Malim, M. H. (2004). Cytidine deamination of retroviral DNA by diverse APOBEC proteins. *Curr. Biol.*, 14(15), 1392-1396.
- Blanc, V., Kennedy, S., & Davidson, N. O. (2003). A novel nuclear localization signal in the auxiliary domain of apobec-1 complementation factor regulates nucleocytoplasmic import and shuttling. *J Biol Chem*, 278(42), 41198-41204.
- Bogerd, H. P., Wiegand, H. L., Hulme, A. E., Garcia-Perez, J. L., O'Shea, K. S., Moran, J. V., & Cullen, B. R. (2006). Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc Natl Acad Sci USA*, 103(23), 8780-8785. doi: 0603313103 [pii]
10.1073/pnas.0603313103
- Bohn, M. F., Shandilya, S. M., Silvas, T. V., Nalivaika, E. A., Kouno, T., Kelch, B. A., . . . Schiffer, C. A. (2015). The ssDNA Mutator APOBEC3A Is Regulated by Cooperative Dimerization. *Structure*, 23(5), 903-911. doi: 10.1016/j.str.2015.03.016
- Bonhomme, D., Hammarstrom, L., Webster, D., Chapel, H., Hermine, O., Le Deist, F., . . . Levy, Y. (2000). Impaired antibody affinity maturation process characterizes a subset of patients with common variable immunodeficiency. *J Immunol*, 165(8), 4725-4730.
- Bonvin, M., Achermann, F., Greeve, I., Stroka, D., Keogh, A., Inderbitzin, D., . . . Greeve, J. (2006). Interferon-inducible expression of APOBEC3 editing enzymes in human hepatocytes and inhibition of hepatitis B virus replication. *Hepatology*, 43(6), 1364-1374.
- Brar, S. S., Sacho, E. J., Tessmer, I., Croteau, D. L., Erie, D. A., & Diaz, M. (2008). Activation-induced deaminase, AID, is catalytically active as a monomer on single-stranded DNA. *DNA Repair (Amst)*, 7(1), 77-87. doi: 10.1016/j.dnarep.2007.08.002
- Burns, M. B., Lackey, L., Carpenter, M. A., Rathore, A., Land, A. M., Leonard, B., . . . Harris, R. S. (2013). APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature*, 494(7437), 366-370. doi: 10.1038/nature11881
- Burns, M. B., Temiz, N. A., & Harris, R. S. (2013). Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat Genet*, 45(9), 977-983. doi: 10.1038/ng.2701
- Byeon, I. J., Ahn, J., Mitra, M., Byeon, C. H., Hercik, K., Hritz, J., . . . Gronenborn, A. M. (2013). NMR structure of human restriction factor APOBEC3A reveals substrate binding and enzyme specificity. *Nat Commun*, 4, 1890. doi: 10.1038/ncomms2883
- Carpenter, M. A., Li, M., Rathore, A., Lackey, L., Law, E. K., Land, A. M., . . . Harris, R. S. (2012). Methylcytosine and Normal Cytosine Deamination by the Foreign DNA

Restriction Enzyme APOBEC3A. *J Biol Chem*, 287(41), 34801-34808. doi: 10.1074/jbc.M112.385161

-C-

- Caval, V., Suspene, R., Shapira, M., Vartanian, J. P., & Wain-Hobson, S. (2014). A prevalent cancer susceptibility APOBEC3A hybrid allele bearing APOBEC3B 3'UTR enhances chromosomal DNA damage. *Nat Commun*, 5, 5129. doi: 10.1038/ncomms6129
- Caval, V., Suspene, R., Vartanian, J. P., & Wain-Hobson, S. (2014). Orthologous mammalian APOBEC3A cytidine deaminases hypermutate nuclear DNA. *Mol Biol Evol*, 31(2), 330-340. doi: 10.1093/molbev/mst195
- Cescon, D. W., Haibe-Kains, B., & Mak, T. W. (2015). APOBEC3B expression in breast cancer reflects cellular proliferation, while a deletion polymorphism is associated with immune activation. *Proc Natl Acad Sci U S A*, 112(9), 2841-2846. doi: 10.1073/pnas.1424869112
- Chaipan, C., Smith, J. L., Hu, W. S., & Pathak, V. K. (2013). APOBEC3G restricts HIV-1 to a greater extent than APOBEC3F and APOBEC3DE in human primary CD4+ T cells and macrophages. *J Virol*, 87(1), 444-453. doi: 10.1128/JVI.00676-12
- Chelico, L., Prochnow, C., Erie, D. A., Chen, X. S., & Goodman, M. F. (2010). Structural model for deoxycytidine deamination mechanisms of the HIV-1 inactivation enzyme APOBEC3G. *J Biol Chem*, 285(21), 16195-16205. doi: [M110.107987 \[pii\]](https://doi.org/10.1074/jbc.M110.107987)
10.1074/jbc.M110.107987
- Chen, H., Lilley, C. E., Yu, Q., Lee, D. V., Chou, J., Narvaiza, I., . . . Weitzman, M. D. (2006). APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. *Curr Biol*, 16(5), 480-485. doi: S0960-9822(06)01040-2 [pii]
10.1016/j.cub.2006.01.031
- Chen, K., Huang, J., Zhang, C., Huang, S., Nunnari, G., Wang, F. X., . . . Zhang, H. (2006). Alpha interferon potently enhances the anti-human immunodeficiency virus type 1 activity of APOBEC3G in resting primary CD4 T cells. *J Virol*, 80(15), 7645-7657. doi: 80/15/7645 [pii]
10.1128/JVI.00206-06
- Chester, A., Somasekaram, A., Tzimina, M., Jarmuz, A., Gisbourne, J., O'Keefe, R., . . . Navaratnam, N. (2003). The apolipoprotein B mRNA editing complex performs a multifunctional cycle and suppresses nonsense-mediated decay. *EMBO J*, 22(15), 3971-3982. doi: 10.1093/emboj/cdg369
- Cheynier, R., Gratton, S., Vartanian, J. P., Meyerhans, A., & Wain-Hobson, S. (1997). G --> A hypermutation does not result from polymerase chain reaction. *AIDS Res Hum Retroviruses*, 13(12), 985-986.
- Chiu, Y. H., Macmillan, J. B., & Chen, Z. J. (2009). RNA polymerase III detects cytosolic DNA and induces type I interferons through the RIG-I pathway. *Cell*, 138(3), 576-591. doi: 10.1016/j.cell.2009.06.015
- Chiu, Y. L., Witkowska, H. E., Hall, S. C., Santiago, M., Soros, V. B., Esnault, C., . . . Greene, W. C. (2006). High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition. *Proc Natl Acad Sci U S A*, 103(42), 15588-15593. doi: 0604524103 [pii]
10.1073/pnas.0604524103
- Chung, S. J., Fromme, J. C., & Verdine, G. L. (2005). Structure of human cytidine deaminase bound to a potent inhibitor. *J Med Chem*, 48(3), 658-660.

- Conticello, S. G., Thomas, C. J., Petersen-Mahrt, S. K., & Neuberger, M. S. (2005). Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol. Biol. Evol.*, 22(2), 367-377.
- Costanzi, S., Vincenzetti, S., Cristalli, G., & Vita, A. (2006). Human cytidine deaminase: a three-dimensional homology model of a tetrameric metallo-enzyme inferred from the crystal structure of a distantly related dimeric homologue. *J Mol Graph Model*, 25(1), 10-16. doi: 10.1016/j.jmglm.2005.10.008

-D-

- Dang, Y., Abudu, A., Son, S., Harjes, E., Spearman, P., Matsuo, H., & Zheng, Y. H. (2011). Identification of a single amino acid required for APOBEC3 antiretroviral cytidine deaminase activity. *J Virol*, 85(11), 5691-5695. doi: 10.1128/JVI.00243-11
- Dang, Y., Wang, X., Esselman, W. J., & Zheng, Y. H. (2006). Identification of APOBEC3DE as another antiretroviral factor from the human APOBEC family. *J Virol*, 80(21), 10522-10533.
- Datta, J., Kutay, H., Nasser, M. W., Nuovo, G. J., Wang, B., Majumder, S., . . . Jacob, S. T. (2008). Methylation mediated silencing of MicroRNA-1 gene and its role in hepatocellular carcinogenesis. *Cancer Res*, 68(13), 5049-5058. doi: 10.1158/0008-5472.CAN-07-6655
- Delebecque, F., Suspene, R., Calattini, S., Casartelli, N., Saib, A., Froment, A., . . . Schwartz, O. (2006). Restriction of foamy viruses by APOBEC cytidine deaminases. *J Virol*, 80(2), 605-614.
- Derse, D., Hill, S. A., Princler, G., Lloyd, P., & Heidecker, G. (2007). Resistance of human T cell leukemia virus type 1 to APOBEC3G restriction is mediated by elements in nucleocapsid. *Proc Natl Acad Sci U S A*, 104(8), 2915-2920. doi: 10.1073/pnas.0609444104
- Ding, L., Getz, G., Wheeler, D. A., Mardis, E. R., McLellan, M. D., Cibulskis, K., . . . Wilson, R. K. (2008). Somatic mutations affect key pathways in lung adenocarcinoma. *Nature*, 455(7216), 1069-1075. doi: nature07423 [pii] 10.1038/nature07423
- Drake, J. W., Charlesworth, B., Charlesworth, D., & Crow, J. F. (1998). Rates of spontaneous mutation. *Genetics*, 148(4), 1667-1686.
- Duggal, N. K., Malik, H. S., & Emerman, M. (2011). The breadth of antiviral activity of APOBEC3DE in chimpanzees has been driven by positive selection. *J Virol*, 85(21), 11361-11371. doi: 10.1128/JVI.05046-11

-E-

- Eckert, K. A., & Kunkel, T. A. (1991). DNA polymerase fidelity and the polymerase chain reaction. *PCR Methods Appl*, 1(1), 17-24.
- Eckert, K. A., & Kunkel, T. A. (1993). Effect of reaction pH on the fidelity and processivity of exonuclease-deficient Klenow polymerase. *J Biol Chem*, 268(18), 13462-13471.
- Esnault, C., Heidmann, O., Delebecque, F., Dewannieux, M., Ribet, D., Hance, A. J., . . . Schwartz, O. (2005). APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. *Nature*, 433(7024), 430-433.
- Esnault, C., Priet, S., Ribet, D., Heidmann, O., & Heidmann, T. (2008). Restriction by APOBEC3 proteins of endogenous retroviruses with an extracellular life cycle: ex vivo effects and in vivo "traces" on the murine IAP and human HERV-K elements. *Retrovirology*, 5, 75. doi: 10.1186/1742-4690-5-75

-F-

- Fan, J., Ma, G., Nosaka, K., Tanabe, J., Satou, Y., Koito, A., . . . Matsuoka, M. (2010). APOBEC3G generates nonsense mutations in human T-cell leukemia virus type 1 proviral genomes in vivo. *J Virol*, 84(14), 7278-7287. doi: JVI.02239-09 [pii] 10.1128/JVI.02239-09
- Fernandes-Alnemri, T., Yu, J. W., Datta, P., Wu, J., & Alnemri, E. S. (2009). AIM2 activates the inflammasome and cell death in response to cytoplasmic DNA. *Nature*, 458(7237), 509-513. doi: nature07710 [pii] 10.1038/nature07710
- Fitting, C., Parlato, M., Adib-Conquy, M., Memain, N., Philippart, F., Misset, B., . . . Adrie, C. (2012). DNAemia detection by multiplex PCR and biomarkers for infection in systemic inflammatory response syndrome patients. *PLoS One*, 7(6), e38916. doi: 10.1371/journal.pone.0038916
- Friew, Y. N., Boyko, V., Hu, W. S., & Pathak, V. K. (2009). Intracellular interactions between APOBEC3G, RNA, and HIV-1 Gag: APOBEC3G multimerization is dependent on its association with RNA. *Retrovirology*, 6, 56. doi: 10.1186/1742-4690-6-56
- Fujino, T., Navaratnam, N., & Scott, J. (1998). Human apolipoprotein B RNA editing deaminase gene (APOBEC1). *Genomics*, 47(2), 266-275.

-G-

- Gagneux, P., Wills, C., Gerloff, U., Tautz, D., Morin, P. A., Boesch, C., . . . Woodruff, D. S. (1999). Mitochondrial sequences show diverse evolutionary histories of African hominoids. *Proceedings of the National Academy of Sciences U.S.A.*, 96(9), 5077-5082.
- Gao, F., Yue, L., White, A. T., Pappas, P. G., Barchue, J., Hanson, A. P., . . . Hahn, B. H. (1992). Human infection by genetically diverse SIVSM-related HIV-2 in west Africa. *Nature*, 358(6386), 495-499. doi: 10.1038/358495a0
- Gonzalez, M. C., Suspene, R., Henry, M., Guetard, D., Wain-Hobson, S., & Vartanian, J. P. (2009). Human APOBEC1 cytidine deaminase edits HBV DNA. *Retrovirology*, 6, 96. doi: 1742-4690-6-96 [pii] 10.1186/1742-4690-6-96
- Goodenow, M., Huet, T., Saurin, W., Kwok, S., Sninsky, J., & Wain-Hobson, S. (1989). HIV-1 isolates are rapidly evolving quasispecies: evidence for viral mixtures and preferred nucleotide substitutions. *J Acquir Immune Defic Syndr*, 2(4), 344-352.
- Goodsell, D. S. (2001). The molecular perspective: ultraviolet light and pyrimidine dimers. *Stem Cells*, 19(4), 348-349. doi: 10.1634/stemcells.19-4-348
- Greenman, C., Stephens, P., Smith, R., Dalgliesh, G. L., Hunter, C., Bignell, G., . . . Stratton, M. R. (2007). Patterns of somatic mutation in human cancer genomes. *Nature*, 446(7132), 153-158. doi: nature05610 [pii] 10.1038/nature05610
- Greeve, J., Altkemper, I., Dieterich, J. H., Greten, H., & Windler, E. (1993). Apolipoprotein B mRNA editing in 12 different mammalian species: hepatic expression is reflected in low concentrations of apoB-containing plasma lipoproteins. *J Lipid Res*, 34(8), 1367-1383.

Günther, S., Sommer, G., Plikat, U., Iwanska, A., Wain-Hobson, S., Will, H., & Meyerhans, A. (1997). Naturally occurring hepatitis B virus genomes bearing the hallmarks of retroviral G→A hypermutation. *Virology*, 235(1), 104-108.

-H-

Hache, G., Liddament, M. T., & Harris, R. S. (2005). The retroviral hypermutation specificity of APOBEC3F and APOBEC3G is governed by the C-terminal DNA cytosine deaminase domain. *J Biol Chem*, 280(12), 10920-10924. doi: 10.1074/jbc.M500382200

Harris, R. S., Bishop, K. N., Sheehy, A. M., Craig, H. M., Petersen-Mahrt, S. K., Watt, I. N., . . . Malim, M. H. (2003). DNA deamination mediates innate immunity to retroviral infection. *Cell*, 113(6), 803-809.

Harris, R. S., & Liddament, M. T. (2004). Retroviral restriction by APOBEC proteins. *Nat Rev Immunol*, 4(11), 868-877.

Harris, R. S., Sale, J. E., Petersen-Mahrt, S. K., & Neuberger, M. S. (2002). AID is essential for immunoglobulin V gene conversion in a cultured B cell line. *Curr Biol*, 12(5), 435-438.

Hazkani-Covo, E., Zeller, R. M., & Martin, W. (2010). Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLoS Genet*, 6(2), e1000834. doi: 10.1371/journal.pgen.1000834

Henry, M., Guetard, D., Suspene, R., Rusniok, C., Wain-Hobson, S., & Vartanian, J. P. (2009). Genetic editing of HBV DNA by monodomain human APOBEC3 cytidine deaminases and the recombinant nature of APOBEC3G. *PLoS One*, 4(1), e4277. doi: 10.1371/journal.pone.0004277

Holmes, R. K., Koning, F. A., Bishop, K. N., & Malim, M. H. (2007). APOBEC3F can inhibit the accumulation of HIV-1 reverse transcription products in the absence of hypermutation. Comparisons with APOBEC3G. *J Biol Chem*, 282(4), 2587-2595. doi: 10.1074/jbc.M607298200

Huang, X., & Yang, Y. (2010). Targeting the TLR9-MyD88 pathway in the regulation of adaptive immune responses. *Expert Opin Ther Targets*, 14(8), 787-796. doi: 10.1517/14728222.2010.501333

Hultquist, J. F., Lengyel, J. A., Refsland, E. W., Larue, R. S., Lackey, L., Brown, W. L., & Harris, R. S. (2011). Human and Rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H Demonstrate a Conserved Capacity to Restrict Vif-deficient HIV-1. *J Virol*, 84(19), 11220-112234. doi: JVI.05238-11 [pii] 10.1128/JVI.05238-11

Huthoff, H., Autore, F., Gallois-Montbrun, S., Fraternali, F., & Malim, M. H. (2009). RNA-dependent oligomerization of APOBEC3G is required for restriction of HIV-1. *PLoS Pathog*, 5(3), e1000330. doi: 10.1371/journal.ppat.1000330

-I-

Imai, K., Slupphaug, G., Lee, W. I., Revy, P., Nonoyama, S., Catalan, N., . . . Durandy, A. (2003). Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. *Nat Immunol*, 4(10), 1023-1028. doi: 10.1038/ni974 ni974 [pii]

-J-

- Jankovic, M., Robbiani, D. F., Dorsett, Y., Eisenreich, T., Xu, Y., Tarakhovsky, A., . . . Nussenzweig, M. C. (2010). Role of the translocation partner in protection against AID-dependent chromosomal translocations. *Proc Natl Acad Sci U S A*, 107(1), 187-192. doi: 10.1073/pnas.0908946107
- Jarmuz, A., Chester, A., Bayliss, J., Gisbourne, J., Dunham, I., Scott, J., & Navaratnam, N. (2002). An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics*, 79(3), 285-296.
- Jena, N. R. (2012). DNA damage by reactive species: Mechanisms, mutation and repair. *J Biosci*, 37(3), 503-517.
- Jha, P., Sinha, S., Kanchan, K., Qidwai, T., Narang, A., Singh, P. K., . . . Habib, S. (2012). Deletion of the APOBEC3B gene strongly impacts susceptibility to falciparum malaria. *Infect Genet Evol*, 12(1), 142-148. doi: 10.1016/j.meegid.2011.11.001
- Johnson, P. R., Hamm, T. E., Goldstein, S., Kitov, S., & Hirsch, V. M. (1991). The genetic fate of molecularly cloned simian immunodeficiency virus in experimentally infected macaques. *Virology*, 185(1), 217-228.
- Ju, Y. S., Tubio, J. M., Mifsud, W., Fu, B., Davies, H. R., Ramakrishna, M., . . . Stratton, M. R. (2015). Frequent somatic transfer of mitochondrial DNA into the nuclear genome of human cancer cells. *Genome Res*, 25(6), 814-824. doi: 10.1101/gr.190470.115

-K-

- Kidd, J. M., Newman, T. L., Tuzun, E., Kaul, R., & Eichler, E. E. (2007). Population stratification of a common APOBEC gene deletion polymorphism. *PLoS Genet*, 3(4), e63. doi: 06-PLGE-RA-0479R3 [pii] 10.1371/journal.pgen.0030063
- Kinomoto, M., Kanno, T., Shimura, M., Ishizaka, Y., Kojima, A., Kurata, T., . . . Tokunaga, K. (2007). All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. *Nucleic Acids Res*, 35(9), 2955-2964.
- Kiss-Toth, E., Bagstaff, S. M., Sung, H. Y., Jozsa, V., Dempsey, C., Caunt, J. C., . . . Dower, S. K. (2004). Human tribbles, a protein family controlling mitogen-activated protein kinase cascades. *J Biol Chem*, 279(41), 42703-42708. doi: 10.1074/jbc.M407732200 M407732200 [pii]
- Kock, J., & Blum, H. E. (2008). Hypermutation of hepatitis B virus genomes by APOBEC3G, APOBEC3C and APOBEC3H. *J Gen Virol*, 89(Pt 5), 1184-1191. doi: 89/5/1184 [pii] 10.1099/vir.0.83507-0
- Komatsu, A., Nagasaki, K., Fujimori, M., Amano, J., & Miki, Y. (2008). Identification of novel deletion polymorphisms in breast cancer. *Int J Oncol*, 33(2), 261-270.
- Koning, F. A., Newman, E. N., Kim, E. Y., Kunstman, K. J., Wolinsky, S. M., & Malim, M. H. (2009). Defining APOBEC3 expression patterns in human tissues and hematopoietic cell subsets. *J Virol*, 83(18), 9474-9485. doi: JVI.01089-09 [pii] 10.1128/JVI.01089-09
- Kostrzak, A., Henry, M., Demoyen, P. L., Wain-Hobson, S., & Vartanian, J. P. (2014). APOBEC3A catabolism of electroporated plasmid DNA in mouse muscle. *Gene Ther*. doi: 10.1038/gt.2014.88
- Kow, Y. W. (2002). Repair of deaminated bases in DNA. *Free Radic Biol Med*, 33(7), 886-893.
- Koyama, T., Arias, J. F., Iwabu, Y., Yokoyama, M., Fujita, H., Sato, H., & Tokunaga, K. (2013). APOBEC3G oligomerization is associated with the inhibition of both Alu and

- LINE-1 retrotransposition. *PLoS One*, 8(12), e84228. doi: 10.1371/journal.pone.0084228
- Kulis, M., & Esteller, M. (2010). DNA methylation and cancer. *Adv Genet*, 70, 27-56. doi: 10.1016/B978-0-12-380866-0.60002-2
- Kunkel, T. A. (2004). DNA replication fidelity. *J Biol Chem*, 279(17), 16895-16898. doi: 10.1074/jbc.R400006200

-L-

- Landry, S., Narvaiza, I., Linfesty, D. C., & Weitzman, M. D. (2011). APOBEC3A can activate the DNA damage response and cause cell-cycle arrest. *EMBO reports*, 12(5), 444-450. doi: 10.1038/embor.2011.46
- Larue, R. S., Andresdottir, V., Blanchard, Y., Conticello, S. G., Derse, D., Emerman, M., . . . Harris, R. S. (2008). Guidelines for Naming Non-Primate APOBEC3 Genes and Proteins. *J Virol*, 83, 494-497. doi: JVI.01976-08 [pii] 10.1128/JVI.01976-08
- LaRue, R. S., Jonsson, S. R., Silverstein, K. A., Lajoie, M., Bertrand, D., El-Mabrouk, N., . . . Harris, R. S. (2008). The artiodactyl APOBEC3 innate immune repertoire shows evidence for a multi-functional domain organization that existed in the ancestor of placental mammals. *BMC Mol Biol*, 9, 104. doi: 10.1186/1471-2199-9-104
- Lecossier, D., Bouchomet, F., Clavel, F., & Hance, A. J. (2003). Hypermutation of HIV-1 DNA in the absence of the Vif protein. *Science*, 300(5622), 1112.
- Li, J., Chen, Y., Li, M., Carpenter, M. A., McDougle, R. M., Luengas, E. M., . . . Mueller, J. D. (2014). APOBEC3 multimerization correlates with HIV-1 packaging and restriction activity in living cells. *J Mol Biol*, 426(6), 1296-1307. doi: 10.1016/j.jmb.2013.12.014
- Liang, G., Kitamura, K., Wang, Z., Liu, G., Chowdhury, S., Fu, W., . . . Muramatsu, M. (2013). RNA editing of hepatitis B virus transcripts by activation-induced cytidine deaminase. *Proc Natl Acad Sci U S A*, 110(6), 2246-2251. doi: 10.1073/pnas.1221921110
- Liao, W., Hong, S. H., Chan, B. H., Rudolph, F. B., Clark, S. C., & Chan, L. (1999). APOBEC-2, a cardiac- and skeletal muscle-specific member of the cytidine deaminase supergene family. *Biochem Biophys Res Commun*, 260(2), 398-404.
- Lochelt, M., Romen, F., Bastone, P., Muckenfuss, H., Kirchner, N., Kim, Y. B., . . . Munk, C. (2005). The antiretroviral activity of APOBEC3 is inhibited by the foamy virus accessory Bet protein. *Proc Natl Acad Sci U S A*, 102(22), 7982-7987. doi: 0501445102 [pii] 10.1073/pnas.0501445102
- Lucifora, J., Xia, Y., Reisinger, F., Zhang, K., Stadler, D., Cheng, X., . . . Protzer, U. (2014). Specific and nonhepatotoxic degradation of nuclear hepatitis B virus cccDNA. *Science*, 343(6176), 1221-1228. doi: 10.1126/science.1243462

-M-

- Madsen, P., Anant, S., Rasmussen, H. H., Gromov, P., Vorum, H., Dumanski, J. P., . . . Celis, J. E. (1999). Psoriasis upregulated phorbol-1 shares structural but not functional similarity to the mRNA-editing protein apobec-1. *J Invest Dermatol*, 113(2), 162-169.
- Mahieux, R., Suspene, R., Delebecque, F., Henry, M., Schwartz, O., Wain-Hobson, S., & Vartanian, J. P. (2005). Extensive editing of a small fraction of human T-cell leukemia

- virus type 1 genomes by four APOBEC3 cytidine deaminases. *J Gen Virol*, 86(Pt 9), 2489-2494.
- Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L., & Trono, D. (2003). Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature*, 424(6944), 99-103.
- Mansson, B., Geborek, P., Saxne, T., & Bjornsson, S. (1990). Cytidine deaminase activity in synovial fluid of patients with rheumatoid arthritis: relation to lactoferrin, acidosis, and cartilage proteoglycan release. *Ann Rheum Dis*, 49(8), 594-597.
- Marchio, A., Pineau, P., Meddeb, M., Terris, B., Tiollais, P., Bernheim, A., & Dejean, A. (2000). Distinct chromosomal abnormality pattern in primary liver cancer of non-B, non-C patients. *Oncogene*, 19(33), 3733-3738.
- Mariani, R., Chen, D., Schrofelbauer, B., Navarro, F., Konig, R., Bollman, B., . . . Landau, N. R. (2003). Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. *Cell*, 114(1), 21-31.
- Marin, M., Rose, K. M., Kozak, S. L., & Kabat, D. (2003). HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. *Nat. Med.*, 9(11), 1398-1403.
- Mayumi-Matsuda, K., Kojima, S., Suzuki, H., & Sakata, T. (1999). Identification of a novel kinase-like gene induced during neuronal cell death. *Biochem Biophys Res Commun*, 258(2), 260-264. doi: 10.1006/bbrc.1999.0576
- McDougall, W. M., Okany, C., & Smith, H. C. (2011). Deaminase activity on single-stranded DNA (ssDNA) occurs in vitro when APOBEC3G cytidine deaminase forms homotetramers and higher-order complexes. *J Biol Chem*, 286(35), 30655-30661. doi: 10.1074/jbc.M111.269506
- Meyerhans, A., Vartanian, J. P., & Wain-Hobson, S. (1990). DNA recombination during PCR. *Nucleic Acids Res*, 18(7), 1687-1691.
- Minegishi, Y., Lavoie, A., Cunningham-Rundles, C., Bedard, P. M., Hebert, J., Cote, L., . . . Conley, M. E. (2000). Mutations in activation-induced cytidine deaminase in patients with hyper IgM syndrome. *Clin Immunol*, 97(3), 203-210. doi: 10.1006/clim.2000.4956
- Miyazaki, Y., Fujinami, M., Inoue, H., Kikuchi, K., Ide, F., & Kusama, K. (2013). Expression of activation-induced cytidine deaminase in oral epithelial dysplasia and oral squamous cell carcinoma. *J Oral Sci*, 55(4), 293-299.
- Miyoshi, Y., Iwao, K., Nagasawa, Y., Aihara, T., Sasaki, Y., Imaoka, S., . . . Nakamura, Y. (1998). Activation of the beta-catenin gene in primary hepatocellular carcinomas by somatic alterations involving exon 3. *Cancer Res*, 58(12), 2524-2527.
- Mohanram, V., Skold, A. E., Bachle, S. M., Pathak, S. K., & Spetz, A. L. (2013). IFN-alpha induces APOBEC3G, F, and A in immature dendritic cells and limits HIV-1 spread to CD4+ T cells. *J Immunol*, 190(7), 3346-3353. doi: 10.4049/jimmunol.1201184
- Montgomery, S. B., Goode, D. L., Kvikstad, E., Albers, C. A., Zhang, Z. D., Mu, X. J., . . . Lunter, G. (2013). The origin, evolution, and functional impact of short insertion-deletion variants identified in 179 human genomes. *Genome Res*, 23(5), 749-761. doi: 10.1101/gr.148718.112
- Moris, A., Murray, S., & Cardinaud, S. (2014). AID and APOBECs span the gap between innate and adaptive immunity. *Front Microbiol*, 5, 534. doi: 10.3389/fmicb.2014.00534
- Muckenfuss, H., Hamdorf, M., Held, U., Perkovic, M., Lower, J., Cichutek, K., . . . Munk, C. (2006). APOBEC3 proteins inhibit human LINE-1 retrotransposition. *J Biol Chem*, 281(31), 22161-22172. doi: M601716200 [pii] 10.1074/jbc.M601716200

- Muckenfuss, H., Kaiser, J. K., Krebil, E., Battenberg, M., Schwer, C., Cichutek, K., . . . Flory, E. (2007). Sp1 and Sp3 regulate basal transcription of the human APOBEC3G gene. *Nucleic Acids Res*, 35(11), 3784-3796. doi: 10.1093/nar/gkm340
- Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., & Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell*, 102(5), 553-563. doi: S0092-8674(00)00078-7 [pii]
- Muramatsu, M., Sankaranand, V. S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N. O., & Honjo, T. (1999). Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J Biol Chem*, 274(26), 18470-18476.
- Muschen, M., Re, D., Jungnickel, B., Diehl, V., Rajewsky, K., & Kuppers, R. (2000). Somatic mutation of the CD95 gene in human B cells as a side-effect of the germinal center reaction. *J Exp Med*, 192(12), 1833-1840.
- Mussil, B., Suspène, R., Aynaud, M. M., Gauvrit, A., Vartanian, J. P., & Wain-Hobson, S. (2013). Human APOBEC3A isoforms translocate to the nucleus and induce DNA double strand breaks leading to cell stress and death. *PLoS One*, 8(8), e73641. doi: 10.1371/journal.pone.0073641

-N-

- Narvaiza, I., Linfesty, D. C., Greener, B. N., Hakata, Y., Pintel, D. J., Logue, E., . . . Weitzman, M. D. (2009). Deaminase-independent inhibition of parvoviruses by the APOBEC3A cytidine deaminase. *PLoS Pathog*, 5(5), e1000439. doi: 10.1371/journal.ppat.1000439
- Newman, E. N., Holmes, R. K., Craig, H. M., Klein, K. C., Lingappa, J. R., Malim, M. H., & Sheehy, A. M. (2005). Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. *Curr Biol*, 15(2), 166-170. doi: S0960982204010498 [pii] 10.1016/j.cub.2004.12.068
- Nguyen, D. H., Gummuluru, S., & Hu, J. (2007). Deamination-independent inhibition of hepatitis B virus reverse transcription by APOBEC3G. *J Virol*, 81(9), 4465-4472. doi: JVI.02510-06 [pii] 10.1128/JVI.02510-06
- Niavarani, A., Currie, E., Rey, Y., Anjos-Afonso, F., Horswell, S., Griessinger, E., . . . Bonnet, D. (2015). APOBEC3A is implicated in a novel class of G-to-A mRNA editing in WT1 transcripts. *PLoS One*, 10(3), e0120089. doi: 10.1371/journal.pone.0120089
- Nik-Zainal, S., Alexandrov, L. B., Wedge, D. C., Van Loo, P., Greenman, C. D., Raine, K., . . . Stratton, M. R. (2012). Mutational Processes Molding the Genomes of 21 Breast Cancers. *Cell*, 149(5), 979-993. doi: 10.1016/j.cell.2012.04.024
- Noguchi, C., Hiraga, N., Mori, N., Tsuge, M., Imamura, M., Takahashi, S., . . . Chayama, K. (2007). Dual effect of APOBEC3G on Hepatitis B virus. *J Gen Virol*, 88(Pt 2), 432-440. doi: 88/2/432 [pii] 10.1099/vir.0.82319-0

-O-

- OhAinle, M., Kerns, J. A., Li, M. M., Malik, H. S., & Emerman, M. (2008). Antiretroelement activity of APOBEC3H was lost twice in recent human evolution. *Cell Host Microbe*, 4(3), 249-259. doi: S1931-3128(08)00223-0 [pii]
10.1016/j.chom.2008.07.005
- OhAinle, M., Kerns, J. A., Malik, H. S., & Emerman, M. (2006). Adaptive evolution and antiviral activity of the conserved mammalian cytidine deaminase APOBEC3H. *J Virol*, 80(8), 3853-3862. doi: 80/8/3853 [pii]
10.1128/JVI.80.8.3853-3862.2006
- Ohsugi, T., & Koito, A. (2007). Human T cell leukemia virus type I is resistant to the antiviral effects of APOBEC3. *J Virol Methods*, 139(1), 93-96.
- Okazaki, I. M., Hiai, H., Kakazu, N., Yamada, S., Muramatsu, M., Kinoshita, K., & Honjo, T. (2003). Constitutive expression of AID leads to tumorigenesis. *J Exp Med*, 197(9), 1173-1181. doi: 10.1084/jem.20030275
jem.20030275 [pii]
- Okuyama, S., Marusawa, H., Matsumoto, T., Ueda, Y., Matsumoto, Y., Endo, Y., . . . Chiba, T. (2012). Excessive activity of apolipoprotein B mRNA editing enzyme catalytic polypeptide 2 (APOBEC2) contributes to liver and lung tumorigenesis. *Int J Cancer*, 130(6), 1294-1301. doi: 10.1002/ijc.26114
- Ooms, M., Krikoni, A., Kress, A. K., Simon, V., & Munk, C. (2012). APOBEC3A, APOBEC3B, and APOBEC3H haplotype 2 restrict human T-lymphotropic virus type 1. *J Virol*, 86(11), 6097-6108. doi: 10.1128/JVI.06570-11

-P-

- Pasqualucci, A. (1998). Experimental and clinical studies about the preemptive analgesia with local anesthetics. Possible reasons of the failure. *Minerva Anesthesiol*, 64(10), 445-457.
- Pasqualucci, L., Neri, A., Baldini, L., Dalla-Favera, R., & Migliazza, A. (2000). BCL-6 mutations are associated with immunoglobulin variable heavy chain mutations in B-cell chronic lymphocytic leukemia. *Cancer Res*, 60(20), 5644-5648.
- Pasqualucci, L., Neumeister, P., Goossens, T., Nanjangud, G., Chaganti, R. S., Kuppers, R., & Dalla-Favera, R. (2001). Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature*, 412(6844), 341-346. doi: 10.1038/35085588
35085588 [pii]
- Patenaude, A. M., & Di Noia, J. M. (2010). The mechanisms regulating the subcellular localization of AID. *Nucleus*, 1(4), 325-331. doi: 10.4161/nucl.1.4.12107
- Peng, G., Greenwell-Wild, T., Nares, S., Jin, W., Lei, K. J., Rangel, Z. G., . . . Wahl, S. M. (2007). Myeloid differentiation and susceptibility to HIV-1 are linked to APOBEC3 expression. *Blood*, 110(1), 393-400. doi: 10.1182/blood-2006-10-051763
- Peng, G., Lei, K. J., Jin, W., Greenwell-Wild, T., & Wahl, S. M. (2006). Induction of APOBEC3 family proteins, a defensive maneuver underlying interferon-induced anti-HIV-1 activity. *J Exp Med*, 203(1), 41-46.
- Perkovic, M., Schmidt, S., Marino, D., Russell, R. A., Stauch, B., Hofmann, H., . . . Munk, C. (2009). Species-specific inhibition of APOBEC3C by the prototype foamy virus protein bet. *J Biol Chem*, 284(9), 5819-5826. doi: 10.1074/jbc.M808853200
- Petersen-Mahrt, S. K., & Neuberger, M. S. (2003). In vitro deamination of cytosine to uracil in single-stranded DNA by apolipoprotein B editing complex catalytic subunit 1 (APOBEC1). *J Biol Chem*, 278(22), 19583-19586.
- Pineau, P., Marchio, A., Battiston, C., Cordina, E., Russo, A., Terris, B., . . . Dejean, A. (2008). Chromosome instability in human hepatocellular carcinoma depends on p53

- status and aflatoxin exposure. *Mutat Res*, 653(1-2), 6-13. doi: 10.1016/j.mrgentox.2008.01.012
- Pineau, P., Nagai, H., Prigent, S., Wei, Y., Gyapay, G., Weissenbach, J., . . . Dejean, A. (1999). Identification of three distinct regions of allelic deletions on the short arm of chromosome 8 in hepatocellular carcinoma. *Oncogene*, 18(20), 3127-3134. doi: 10.1038/sj.onc.1202648
- Pleasance, E. D., Cheetham, R. K., Stephens, P. J., McBride, D. J., Humphray, S. J., Greenman, C. D., . . . Stratton, M. R. (2010). A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature*, 463(7278), 191-196. doi: 10.1038/nature08658
- Polson, A. G., Bass, B. L., & Casey, J. L. (1996). RNA editing of hepatitis delta virus antigenome by dsRNA-adenosine deaminase. *Nature*, 380(6573), 454-456. doi: 10.1038/380454a0
- Powell, L. M., Wallis, S. C., Pease, R. J., Edwards, Y. H., Knott, T. J., & Scott, J. (1987). A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. *Cell*, 50(6), 831-840.
- Prochnow, C., Bransteitter, R., Klein, M. G., Goodman, M. F., & Chen, X. S. (2007). The APOBEC-2 crystal structure and functional implications for the deaminase AID. *Nature*, 445(7126), 447-451.
- R-
- Rai, K., Huggins, I. J., James, S. R., Karpf, A. R., Jones, D. A., & Cairns, B. R. (2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell*, 135(7), 1201-1212. doi: 10.1016/j.cell.2008.11.042
- Refsland, E. W., Stenglein, M. D., Shindo, K., Albin, J. S., Brown, W. L., & Harris, R. S. (2010). Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction. *Nucleic Acids Res*, 38, 4274-4284. doi: [gkq174 \[pii\]](#)
[10.1093/nar/gkq174](#)
- Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., . . . Durandy, A. (2000). Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HGM2). *Cell*, 102(5), 565-575.
- Rogozin, I. B., Basu, M. K., Jordan, I. K., Pavlov, Y. I., & Koonin, E. V. (2005). APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases predicted by computational analysis. *Cell Cycle*, 4(9), 1281-1285. doi: 1994 [pii]
- Rosenberg, B. R., Hamilton, C. E., Mwangi, M. M., Dewell, S., & Papavasiliou, F. N. (2011). Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3' UTRs. *Nat Struct Mol Biol*, 18(2), 230-236. doi: 10.1038/nsmb.1975
- Rosler, C., Kock, J., Kann, M., Malim, M. H., Blum, H. E., Baumert, T. F., & von Weizsacker, F. (2005). APOBEC-mediated interference with hepadnavirus production. *Hepatology*, 42(2), 301-309. doi: 10.1002/hep.20801

Russell, R. A., Wiegand, H. L., Moore, M. D., Schafer, A., McClure, M. O., & Cullen, B. R. (2005). Foamy virus Bet proteins function as novel inhibitors of the APOBEC3 family of innate antiretroviral defense factors. *J Virol*, 79(14), 8724-8731.

-S-

- Saccone, S., Besati, C., Andreozzi, L., Della Valle, G., Garattini, E., & Terao, M. (1994). Assignment of the human cytidine deaminase (CDA) gene to chromosome 1 band p35-p36.2. *Genomics*, 22(3), 661-662. doi: 10.1006/geno.1994.1448
- Sakai, S., Ohoka, N., Onozaki, K., Kitagawa, M., Nakanishi, M., & Hayashi, H. (2010). Dual mode of regulation of cell division cycle 25 A protein by TRB3. *Biol Pharm Bull*, 33(7), 1112-1116. doi: JST.JSTAGE/bpb/33.1112 [pii]
- Sakurai, D., Iwatani, Y., Ohtani, H., Naruse, T. K., Terunuma, H., Sugiura, W., & Kimura, A. (2015). APOBEC3H polymorphisms associated with the susceptibility to HIV-1 infection and AIDS progression in Japanese. *Immunogenetics*, 67(4), 253-257. doi: 10.1007/s00251-015-0829-2
- Salter, J. D., Krucinska, J., Raina, J., Smith, H. C., & Wedekind, J. E. (2009). A hydrodynamic analysis of APOBEC3G reveals a monomer-dimer-tetramer self-association that has implications for anti-HIV function. *Biochemistry*, 48(45), 10685-10687. doi: 10.1021/bi901642c
- Santivasi, W. L., & Xia, F. (2014). Ionizing radiation-induced DNA damage, response, and repair. *Antioxid Redox Signal*, 21(2), 251-259. doi: 10.1089/ars.2013.5668
- Sasada, A., Takaori-Kondo, A., Shirakawa, K., Kobayashi, M., Abudu, A., Hishizawa, M., . . . Uchiyama, T. (2005). APOBEC3G targets human T-cell leukemia virus type 1. *Retrovirology*, 2, 32.
- Sato, K., Takeuchi, J. S., Misawa, N., Izumi, T., Kobayashi, T., Kimura, Y., . . . Koyanagi, Y. (2014). APOBEC3D and APOBEC3F potently promote HIV-1 diversification and evolution in humanized mouse model. *PLoS Pathog*, 10(10), e1004453. doi: 10.1371/journal.ppat.1004453
- Sato, Y., Probst, H. C., Tatsumi, R., Ikeuchi, Y., Neuberger, M. S., & Rada, C. (2010). Deficiency in APOBEC2 leads to a shift in muscle fiber type, diminished body mass, and myopathy. *J Biol Chem*, 285(10), 7111-7118. doi: 10.1074/jbc.M109.052977
- Sawyer, S. L., Emerman, M., & Malik, H. S. (2004). Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS Biol*, 2(9), E275. doi: 10.1371/journal.pbio.0020275
- Schoggins, J. W., Wilson, S. J., Panis, M., Murphy, M. Y., Jones, C. T., Bieniasz, P., & Rice, C. M. (2011). A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature*, 472(7344), 481-485. doi: 10.1038/nature09907
- Schulte, W., Bernhagen, J., & Bucala, R. (2013). Cytokines in sepsis: potent immunoregulators and potential therapeutic targets--an updated view. *Mediators Inflamm*, 2013, 165974. doi: 10.1155/2013/165974
- Schumacher, A. J., Nissley, D. V., & Harris, R. S. (2005). APOBEC3G hypermutates genomic DNA and inhibits Ty1 retrotransposition in yeast. *Proc Natl Acad Sci U S A*, 102(28), 9854-9859. doi: 0501694102 [pii]
- 10.1073/pnas.0501694102
- Schumann, G. G. (2007). APOBEC3 proteins: major players in intracellular defence against LINE-1-mediated retrotransposition. *Biochem Soc Trans*, 35(Pt 3), 637-642. doi: 10.1042/BST0350637

- Segasothy, M., & Phillips, P. A. (2001). Systemic lupus erythematosus in Aborigines and Caucasians in central Australia: a comparative study. *Lupus*, 10(6), 439-444.
- Serdjebi, C., Milano, G., & Ciccolini, J. (2015). Role of cytidine deaminase in toxicity and efficacy of nucleosidic analogs. *Expert Opin Drug Metab Toxicol*, 11(5), 665-672. doi: 10.1517/17425255.2015.985648
- Sharma, S., Patnaik, S. K., Taggart, R. T., Kannisto, E. D., Enriquez, S. M., Gollnick, P., & Baysal, B. E. (2015). APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. *Nat Commun*, 6, 6881. doi: 10.1038/ncomms7881
- Sheehy, A. M., Gaddis, N. C., Choi, J. D., & Malim, M. H. (2002). Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature*, 418(6898), 646-650.
- Sheehy, A. M., Gaddis, N. C., & Malim, M. H. (2003). The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nat. Med.*, 9(11), 1404-1407.
- Shen, J. C., Rideout, W. M., 3rd, & Jones, P. A. (1994). The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic Acids Res*, 22(6), 972-976.
- Shimizu, T., Marusawa, H., Matsumoto, Y., Inuzuka, T., Ikeda, A., Fujii, Y., . . . Chiba, T. (2014). Accumulation of somatic mutations in TP53 in gastric epithelium with *Helicobacter pylori* infection. *Gastroenterology*, 147(2), 407-417 e403. doi: 10.1053/j.gastro.2014.04.036
- Shinmura, K., Igarashi, H., Goto, M., Tao, H., Yamada, H., Matsuura, S., . . . Sugimura, H. (2011). Aberrant expression and mutation-inducing activity of AID in human lung cancer. *Ann Surg Oncol*, 18(7), 2084-2092. doi: 10.1245/s10434-011-1568-8
- Shinohara, M., Io, K., Shindo, K., Matsui, M., Sakamoto, T., Tada, K., . . . Takaori-Kondo, A. (2012). APOBEC3B can impair genomic stability by inducing base substitutions in genomic DNA in human cells. *Sci Rep*, 2, 806. doi: 10.1038/srep00806
- Shlyakhtenko, L. S., Lushnikov, A. Y., Li, M., Lackey, L., Harris, R. S., & Lyubchenko, Y. L. (2011). Atomic force microscopy studies provide direct evidence for dimerization of the HIV restriction factor APOBEC3G. *J Biol Chem*, 286(5), 3387-3395. doi: 10.1074/jbc.M110.195685
- Skuse, G. R., Cappione, A. J., Sowden, M., Metheny, L. J., & Smith, H. C. (1996). The neurofibromatosis type I messenger RNA undergoes base-modification RNA editing. *Nucleic Acids Res*, 24(3), 478-485.
- Smith, H. C., Bennett, R. P., Kizilyer, A., McDougall, W. M., & Prohaska, K. M. (2012). Functions and regulation of the APOBEC family of proteins. *Semin Cell Dev Biol*, 23(3), 258-268. doi: 10.1016/j.semcdb.2011.10.004
- Stauch, B., Hofmann, H., Perkovic, M., Weisel, M., Kopietz, F., Cichutek, K., . . . Schneider, G. (2009). Model structure of APOBEC3C reveals a binding pocket modulating ribonucleic acid interaction required for encapsidation. *Proc Natl Acad Sci U S A*, 106(29), 12079-12084. doi: 10.1073/pnas.0900979106
- Stavrou, S., Blouch, K., Kotla, S., Bass, A., & Ross, S. R. (2015). Nucleic acid recognition orchestrates the anti-viral response to retroviruses. *Cell Host Microbe*, 17(4), 478-488. doi: 10.1016/j.chom.2015.02.021
- Stenglein, M. D., Burns, M. B., Li, M., Lengyel, J., & Harris, R. S. (2010). APOBEC3 proteins mediate the clearance of foreign DNA from human cells. *Nat Struct Mol Biol*, 17(2), 222-229. doi: [nsmb.1744](https://doi.org/10.1038/nsmb.1744) [pii] [10.1038/nsmb.1744](https://doi.org/10.1038/nsmb.1744)
- Stephens, P. J., Tarpey, P. S., Davies, H., Van Loo, P., Greenman, C., Wedge, D. C., . . . Stratton, M. R. (2012). The landscape of cancer genes and mutational processes in breast cancer. *Nature*, 486(7403), 400-404. doi: 10.1038/nature11017

- Stopak, K., de Noronha, C., Yonemoto, W., & Greene, W. C. (2003). HIV-1 Vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability. *Mol. Cell.*, 12(3), 591-601.
- Suspène, R., Aynaud, M., Guétard, D., Henry, M., Eckhoff, G., Marchio, A., . . . Wain-Hobson, S. (2011). Somatic hypermutation of human mitochondrial and nuclear DNA by APOBEC3 cytidine deaminases, a pathway for DNA catabolism. *Proc Natl Acad Sci USA*, 108(12), 4858-4863.
- Suspène, R., Aynaud, M. M., Koch, S., Pasdeloup, D., Labetoulle, M., Gaertner, B., . . . Wain-Hobson, S. (2011). Genetic editing of herpes simplex virus 1 and Epstein-Barr herpesvirus genomes by human APOBEC3 cytidine deaminases in culture and in vivo. *J Virol*, 85(15), 7594-7602. doi: JVI.00290-11 [pii]
- 10.1128/JVI.00290-11
- Suspène, R., Aynaud, M. M., Vartanian, J. P., & Wain-Hobson, S. (2013). Efficient deamination of 5-methylcytidine and 5-substituted cytidine residues in DNA by human APOBEC3A cytidine deaminase. *PLoS One*, 8(6), e63461. doi: 10.1371/journal.pone.0063461
- Suspène, R., Guétard, D., Henry, M., Sommer, P., Wain-Hobson, S., & Vartanian, J. P. (2005). Extensive editing of both hepatitis B virus DNA strands by APOBEC3 cytidine deaminases in vitro and in vivo. *Proc Natl Acad Sci USA*, 102(23), 8321-8326.
- Suspène, R., Henry, M., Guillot, S., Wain-Hobson, S., & Vartanian, J. P. (2005). Recovery of APOBEC3-edited human immunodeficiency virus G->A hypermutants by differential DNA denaturation PCR. *J. Gen. Virol.*, 86(Pt 1), 125-129.
- Suspène, R., Sommer, P., Henry, M., Ferris, S., Guétard, D., Pochet, S., . . . Vartanian, J. P. (2004). APOBEC3G is a single-stranded DNA cytidine deaminase and functions independently of HIV reverse transcriptase. *Nucleic Acids Res.*, 32(8), 2421-2429.
- Tanaka, Y., Marusawa, H., Seno, H., Matsumoto, Y., Ueda, Y., Kodama, Y., . . . Chiba, T. (2006). Anti-viral protein APOBEC3G is induced by interferon-alpha stimulation in human hepatocytes. *Biochem Biophys Res Commun*, 341(2), 314-319.
- Teng, B., Burant, C. F., & Davidson, N. O. (1993). Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science*, 260(5115), 1816-1819.
- Teng, B. B., Ochsner, S., Zhang, Q., Soman, K. V., Lau, P. P., & Chan, L. (1999). Mutational analysis of apolipoprotein B mRNA editing enzyme (APOBEC1). structure-function relationships of RNA editing and dimerization. *J Lipid Res*, 40(4), 623-635.

-T-

- Thielen, B. K., McNevin, J. P., McElrath, M. J., Hunt, B. V., Klein, K. C., & Lingappa, J. R. (2010). Innate immune signaling induces high levels of TC-specific deaminase activity in primary monocyte-derived cells through expression of APOBEC3A isoforms. *J Biol Chem*, 285(36), 27753-27766. doi: M110.102822 [pii]
- 10.1074/jbc.M110.102822
- Tobollik, S., Meyer, L., Buettner, M., Klemmer, S., Kempkes, B., Kremmer, E., . . . Jungnickel, B. (2006). Epstein-Barr virus nuclear antigen 2 inhibits AID expression during EBV-driven B-cell growth. *Blood*, 108(12), 3859-3864. doi: 10.1182/blood-2006-05-021303
- Trivedi, N. R., Gilliland, K. L., Zhao, W., Liu, W., & Thiboutot, D. M. (2006). Gene array expression profiling in acne lesions reveals marked upregulation of genes involved in inflammation and matrix remodeling. *J Invest Dermatol*, 126(5), 1071-1079. doi: 10.1038/sj.jid.5700213

Tsuge, M., Noguchi, C., Akiyama, R., Matsushita, M., Kunihiro, K., Tanaka, S., . . . Chayama, K. (2010). G to A hypermutation of TT virus. *Virus Res*(149), 211-216. doi: [10.1016/j.virusres.2010.01.019](https://doi.org/10.1016/j.virusres.2010.01.019)
[S0168-1702\(10\)00045-6 \[pii\]](https://doi.org/10.1016/j.virusres.2010.01.019)

Turelli, P., Mangeat, B., Jost, S., Vianin, S., & Trono, D. (2004). Inhibition of hepatitis B virus replication by APOBEC3G. *Science*, 303(5665), 1829.

-V-

Vartanian, J. P., Guétard, D., Henry, M., & Wain-Hobson, S. (2008). Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. *Science*, 320(5873), 230-233. doi: [10.1126/science.1153201](https://doi.org/10.1126/science.1153201)

Vartanian, J. P., Henry, M., Marchio, A., Suspène, R., Aynaud, M. M., Guétard, D., . . . Wain-Hobson, S. (2010). Massive APOBEC3 editing of hepatitis B viral DNA in cirrhosis. *PLoS Pathog*, 6(5), e1000928. doi: [10.1371/journal.ppat.1000928](https://doi.org/10.1371/journal.ppat.1000928)

Vartanian, J. P., Meyerhans, A., Asjo, B., & Wain-Hobson, S. (1991). Selection, recombination, and G->A hypermutation of human immunodeficiency virus type 1 genomes. *J Virol*, 65(4), 1779-1788.

Vonica, A., Rosa, A., Arduini, B. L., & Brivanlou, A. H. (2011). APOBEC2, a selective inhibitor of TGFbeta signaling, regulates left-right axis specification during early embryogenesis. *Dev Biol*, 350(1), 13-23. doi: [10.1016/j.ydbio.2010.09.016](https://doi.org/10.1016/j.ydbio.2010.09.016)

-W-

Wain-Hobson, S., Sonigo, P., Guyader, M., Gazit, A., & Henry, M. (1995). Erratic G-->A hypermutation within a complete caprine arthritis-encephalitis virus (CAEV) provirus. *Virology*, 209(2), 297-303.

Wang, F. X., Huang, J., Zhang, H., & Ma, X. (2008). APOBEC3G upregulation by alpha interferon restricts human immunodeficiency virus type 1 infection in human peripheral plasmacytoid dendritic cells. *J Gen Virol*, 89(Pt 3), 722-730. doi: [89/3/722 \[pii\]](https://doi.org/10.1099/vir.0.83530-0)
[10.1099/vir.0.83530-0](https://doi.org/10.1099/vir.0.83530-0)

Wang, K., Li, Y., Dai, C., Wang, K., Yu, J., Tan, Y., . . . Yu, X. F. (2013). Characterization of the relationship between APOBEC3B deletion and ACE Alu insertion. *PLoS One*, 8(5), e64809. doi: [10.1371/journal.pone.0064809](https://doi.org/10.1371/journal.pone.0064809)

Wang, Y., Wang, X., Li, J., Zhou, Y., & Ho, W. (2013). RIG-I activation inhibits HIV replication in macrophages. *J Leukoc Biol*, 94(2), 337-341. doi: [10.1189/jlb.0313158](https://doi.org/10.1189/jlb.0313158)

Wang, Z., & Mosbaugh, D. W. (1988). Uracil-DNA glycosylase inhibitor of bacteriophage PBS2: cloning and effects of expression of the inhibitor gene in Escherichia coli. *J Bacteriol*, 170(3), 1082-1091.

Wang, Z., Wakae, K., Kitamura, K., Aoyama, S., Liu, G., Koura, M., . . . Muramatsu, M. (2014). APOBEC3 deaminases induce hypermutation in human papillomavirus 16 DNA upon beta interferon stimulation. *J Virol*, 88(2), 1308-1317. doi: [10.1128/JVI.03091-13](https://doi.org/10.1128/JVI.03091-13)

Warren, C. J., Xu, T., Guo, K., Griffin, L. M., Westrich, J. A., Lee, D., . . . Pyeon, D. (2014). APOBEC3A Functions as a Restriction Factor of Human Papillomavirus. *J Virol*. doi: [10.1128/JVI.02383-14](https://doi.org/10.1128/JVI.02383-14)

- Warren, C. J., Xu, T., Guo, K., Griffin, L. M., Westrich, J. A., Lee, D., . . . Pyeon, D. (2015). APOBEC3A functions as a restriction factor of human papillomavirus. *J Virol*, 89(1), 688-702. doi: 10.1128/JVI.02383-14
- Wedekind, J. E., Gillilan, R., Janda, A., Krucinska, J., Salter, J. D., Bennett, R. P., . . . Smith, H. C. (2006). Nanostructures of APOBEC3G support a hierarchical assembly model of high molecular mass ribonucleoprotein particles from dimeric subunits. *J Biol Chem*, 281(50), 38122-38126. doi: 10.1074/jbc.C600253200
- West, A. P., Khoury-Hanold, W., Staron, M., Tal, M. C., Pineda, C. M., Lang, S. M., . . . Shadel, G. S. (2015). Mitochondrial DNA stress primes the antiviral innate immune response. *Nature*. doi: 10.1038/nature14156
- Wiegand, H. L., Doehle, B. P., Bogerd, H. P., & Cullen, B. R. (2004). A second human antiretroviral factor, APOBEC3F, is suppressed by the HIV-1 and HIV-2 Vif proteins. *EMBO J*, 23(12), 2451-2458.
- Wijesinghe, P., & Bhagwat, A. S. (2012). Efficient deamination of 5-methylcytosines in DNA by human APOBEC3A, but not by AID or APOBEC3G. *Nucleic Acids Res*, 40(18), 9206-9217. doi: 10.1093/nar/gks685
- Wu, J., & Chen, Z. J. (2014). Innate immune sensing and signaling of cytosolic nucleic acids. *Annu Rev Immunol*, 32, 461-488. doi: 10.1146/annurev-immunol-032713-120156
- Wu, L., Pan, J., Thoroddsen, V., Wysong, D. R., Blackman, R. K., Bulawa, C. E., . . . McGovern, K. (2003). Novel small-molecule inhibitors of RNA polymerase III. *Eukaryot Cell*, 2(2), 256-264.

-X-

- Xu, J., Lv, S., Qin, Y., Shu, F., Xu, Y., Chen, J., . . . Wu, J. (2007). TRB3 interacts with CtIP and is overexpressed in certain cancers. *Biochim Biophys Acta*, 1770(2), 273-278. doi: S0304-4165(06)00298-4 [pii]
10.1016/j.bbagen.2006.09.025
- Xu, R., Zhang, X., Zhang, W., Fang, Y., Zheng, S., & Yu, X. F. (2007). Association of human APOBEC3 cytidine deaminases with the generation of hepatitis virus B x antigen mutants and hepatocellular carcinoma. *Hepatology*, 46(6), 1810-1820. doi: 10.1002/hep.21893
- Xuan, D., Li, G., Cai, Q., Deming-Halverson, S., Shrubsole, M. J., Shu, X. O., . . . Long, J. (2013). APOBEC3 deletion polymorphism is associated with breast cancer risk among women of European ancestry. *Carcinogenesis*, 34(10), 2240-2243. doi: 10.1093/carcin/bgt185

-Y-

- Yamanaka, S., Balestra, M. E., Ferrell, L. D., Fan, J., Arnold, K. S., Taylor, S., . . . Innerarity, T. L. (1995). Apolipoprotein B mRNA-editing protein induces hepatocellular carcinoma and dysplasia in transgenic animals. *Proc Natl Acad Sci USA*, 92(18), 8483-8487.
- Yang, P., An, H., Liu, X., Wen, M., Zheng, Y., Rui, Y., & Cao, X. (2010). The cytosolic nucleic acid sensor LRRFIP1 mediates the production of type I interferon via a beta-catenin-dependent pathway. *Nat Immunol*, 11(6), 487-494. doi: 10.1038/ni.1876

Yang, Y., Sowden, M. P., Yang, Y., & Smith, H. C. (2001). Intracellular trafficking determinants in APOBEC-1, the catalytic subunit for cytidine to uridine editing of apolipoprotein B mRNA. *Exp Cell Res*, 267(2), 153-164. doi: 10.1006/excr.2001.5255

-Z-

Zarembek, K. A., & Godowski, P. J. (2002). Tissue expression of human Toll-like receptors and differential regulation of Toll-like receptor mRNAs in leukocytes in response to microbes, their products, and cytokines. *J Immunol*, 168(2), 554-561.

Zhang, T., Cai, J., Chang, J., Yu, D., Wu, C., Yan, T., . . . Lin, D. (2012). Evidence of associations of APOBEC3B gene deletion with susceptibility to persistent HBV infection and hepatocellular carcinoma. *Hum Mol Genet*, 22(6), 1262-1269. doi: 10.1093/hmg/dd513

Zhang, W., Zhang, X., Tian, C., Wang, T., Sarkis, P. T., Fang, Y., . . . Xu, R. (2008). Cytidine deaminase APOBEC3B interacts with heterogeneous nuclear ribonucleoprotein K and suppresses hepatitis B virus expression. *Cell Microbiol*, 10(1), 112-121. doi: 10.1111/j.1462-5822.2007.01020.x

Zielonka, J., Bravo, I. G., Marino, D., Conrad, E., Perkovic, M., Battenberg, M., . . . Munk, C. (2009). Restriction of equine infectious anemia virus by equine APOBEC3 cytidine deaminases. *J Virol*, 83(15), 7547-7559. doi: 10.1128/JVI.00015-09

ANNEXES

1 PLoS Pathogens, 6 Figs, 1S Table

2 SOUMIS

3 APOBEC3DE antagonizes Hepatitis B Virus restriction factors

4 APOBEC3F and APOBEC3G

6 Mohamed S. Bouzidi¹, Vincent Caval¹, Rodolphe Suspène¹, Pascal Pineau²

7 Simon Wain-Hobson¹ and Jean-Pierre Vartanian^{1*}

9 1 Molecular Retrovirology Unit, Institut Pasteur

10 CNRS URA 3015, 28 rue du Dr. Roux,

11 75724 Paris cedex 15, France

13 2 Nuclear Organization and Oncogenesis Unit, Institut Pasteur

14 INSERM U579, 28 rue du Dr. Roux,

15 75724 Paris cedex 15, France

17 *Correspondence to: JPV, Molecular Retrovirology Unit, Institut Pasteur, 28 rue du Dr. Roux,

18 75724 Paris cedex 15, France. T +33 1 44 38 94 45; F +33 1 45 68 88 74;

19 email: jean-pierre.vartanian@pasteur.fr

21 Running title: APOBEC3DE modulates HBV mutagenesis

23 Key words: APOBEC3DE, APOBEC3F, APOBEC3G, HBV, dimerization

24 **Abstract**

25 The *APOBEC3* locus consists of seven genes (*A3A-A3C*, *A3DE*, *A3F-A3H*) that encode DNA
26 cytidine deaminases. These enzymes deaminate single strand DNA, the result being DNA
27 peppered with CG->TA mutations preferentially in the context of TpC with the exception of
28 APOBEC3G (A3G), which prefers CpC dinucleotides. Hepatitis B virus (HBV) is vulnerable
29 to genetic editing of single stranded DNA by APOBEC3 cytidine deaminases, A3G being a
30 major restriction factor for HBV replication. APOBEC3DE (A3DE) stands out in that it is
31 catalytically inactive due to a fixed Tyr320Cys substitution in the C-terminal domain. As
32 A3DE is closely related to A3F and A3G, which can form homo- and heterodimers the impact
33 of A3DE on HBV replication via modulation of other APOBEC3 restriction factors was
34 investigated. A3DE forms dimers with itself, A3F and A3G and antagonizes A3F and to a
35 lesser extent A3G restriction of HBV replication. Ironically while being part of a cluster of
36 innate restriction factors, the A3DE phenotype is pro-viral. As the gorilla genome encodes the
37 same Tyr320Cys substitution this pro-viral phenotype seems to have been selected for.

38
39 **Author summary**

40 The *APOBEC3* locus consists of seven genes that encode DNA cytidine deaminases localized
41 on chromosome 22. These enzymes deaminate single strand DNA, preferentially in the
42 context of TpC. Retroviruses and hepadnaviruses such as hepatitis B virus (HBV) are
43 vulnerable to the negative effects of the APOBEC3F and 3G DNA hypermutator enzymes. It
44 turns out that the APOBEC3DE protein antagonizes the negative effects of these two mutator
45 enzymes so helping HBV to replicate.

Introduction

In the unending battle with microbes, the human genome has evolved a diverse range of innate restriction factors in addition to interferon stimulated genes (ISGs). The *APOBEC3* (*A3*) locus consists of seven genes (*A3A-A3C*, *A3DE*, *A3F-A3H*) (*A3*) that encode DNA cytidine deaminases [1] and arose by duplication and subsequent expansion of the activation-induced deaminase (*AID*) gene in placental mammals [2]. *A3* enzymes impact replication of major viruses such as HIV [3-7], HBV [8-12], herpesviruses [13] and papillomaviruses [14-16] as well as several retroelements [17-19]. They preferentially deaminate single strand DNA (ssDNA), resulting DNA peppered with CG->TA changes, the process being referred to as genetic editing as the changes occur post replication [3-6, 20]. *APOBEC3A* (*A3A*) and *APOBEC3G* (*A3G*) are also ISGs [21-25]. In addition to editing viral DNA, five *A3* enzymes are able to catabolize cytoplasmic mitochondrial DNA, while *A3A* and *APOBEC3B* (*A3B*) can target chromosomal DNA [26-30]. While both enzymes can deaminate 5-methylcytidine [27, 31-33] only *A3A* is able to induce double strand DNA breaks [28, 34, 35]. *A3* enzymes leave an editing signature in DNA: they preferentially edit a C residue in the context of 5'TpC [30, 36] with the exception of *A3G*, which prefers 5'CpC dinucleotides [20, 37].

The enzymes are either made up of a zinc finger single strand DNA binding domain, which is an integral part of the catalytic site (*A3A*, *A3C* and *A3H*), or two such domains with an N-terminal non-catalytic zinc finger domain and a C-terminal catalytic domain (*A3B*, *A3DE*, *A3F* and *A3G*) [1, 38, 39]. While inactive, the N-terminal domain may increase the efficiency of the enzyme [40], many *A3* enzymes can homo- and heterodimerize [41-46] even though *AID*, *A3A* and *A3G* are active *in vitro* as monomers [21, 47, 48]. *A3G*, one of the main restriction factors against human immunodeficiency virus type 1 (HIV-1) and HBV, can also form high molecular weight complexes [21, 47-51].

74 Against such a background, A3DE stands out in that it is catalytically poorly active due to a
75 fixed Tyr320Cys substitution in the C-terminal domain. Mutagenesis to Tyr320 completely
76 restores editing activity against HIV-1 [52]. A non-functional A3DE does not preclude some
77 effect on retroelements or viruses for editing independent restriction has been described for
78 several catalytically inactive mutants [39, 53-55]. Like A3F and A3G, A3DE is a strictly
79 cytoplasmic protein with the double zinc finger domain structure that is also a feature of A3B,
80 which is strictly nuclear. Phylogenetically, the A3DE N-terminal domain (A3D) clusters with
81 that of A3B, A3F and A3G, i.e. [[[A3D, A3Fn]A3Bn]A3Gn], while the C-domain (A3E)
82 groups with the C-terminal domain of A3F and A3C, i.e. [[A3E,A3Fc]A3C] [2, 40, 56].

83

84 A3G represent the major restriction factor for HBV replication with up to 35% of HBV
85 genomes being edited *in vivo* [12]. A3DE, A3F and A3G are co-expressed in a wide range of
86 human tissues [23, 57, 58] and significantly up-regulated in virus-associated cirrhotic liver
87 [12] which was a little surprising for A3DE given that the protein is not functional [57, 59,
88 60]. As A3B, A3F and A3G can form homo- and heterodimers, the impact of A3DE on HBV
89 replication via modulation of other A3 restriction factors was investigated. We show here that
90 A3DE antagonizes the activity of the A3F restriction factor and to a lesser extent A3G
91 resulting in increased HBV replication.

92

93

94 **Results and Discussion**

95 **APOBEC3DE dimerizes with APOBEC3F and APOBEC3G**

96 To investigate A3DE homo- and hetero-dimerization with other A3 enzymes, we performed
97 co-affinity GST pull down experiments. As shown in Fig 1, A3DE interacts strongly with
98 itself, A3F and A3G but failed to pull down A3A, A3B, A3C and A3H. The negative result

for A3B was a little surprising because the N-terminal domains of A3DE and A3B share 79% identical at the protein level, greater than for A3DE and A3G that do interact (43%). To explore this, we used a functional A3B/A chimera construct (A3BnA) encoding the A3B N-terminal domain along with the A3A coding sequences replacing those of the A3B C-terminal which differ by only 9% at the protein level. Now A3DE was able to easily pull down this functional A3BnA hybrid (Fig 1).

In order to confirm the GST pull-down interactions, the intracellular localization of A3-V5 proteins with an N-terminal 3x-FLAG tagged A3DE construct was analyzed by confocal microscopy (Fig 2). A strong and significant co-localization of A3DE with itself, A3F and A3G was observed but not with A3B (Fig 2) and A3A, A3C and A3H (data not shown). However, when the A3BnA construct was used, co-localization between it and A3DE was evident (Fig 2).

The negative findings for A3B, A3A and positive finding for A3BnA show that the heterodimerization with A3DE is mediated by the A3B N-terminal domain while sequences in A3B C-terminal contribute to blocking dimerization. As several residues have been described to be crucial for A3G dimerization [51, 61-63], we made analogous mutants in the A3D domain of A3DE. Tyr28Ala, Trp94Ala, Tyr124Ala and Tyr126Ala mutants, all failed to block A3DE/F or A3DE/G heterodimer formation (data not shown).

As the inactivating A3DE Cys320 site maps away from the catalytic site yet close to residues involved in ssDNA substrate binding, we tested heterodimer formation of the A3DE_{C320Y} mutant with A3F and A3G. By GST-pulldowns A3DE_{C320Y} interacted with itself, wild type A3DE_{C320}, A3F and A3G (Fig 3A). By confocal microscopy the A3DE_{C320Y} mutant behaved

like A3DE and interacted with itself, A3F and A3G (Fig 3B) indicating that functional inactivation of A3DE was unrelated to heterodimer formation or failure to engage with A3B.

APOBEC3DE inhibits APOBEC3F and APOBEC3G editing of HBV DNA

To demonstrate the impact of A3DE/A3F and A3DE/A3G heterodimerization on HBV DNA synthesis, QT6 were co-transfected with pCayw (a molecular clone of the HBV) and A3F or A3G with increasing amounts of A3DE. QT6 cells were used as they do not give an endogenous DNA editing background [12, 30, 40]. Total DNA was recovered at 48 hours and analyzed by 3DPCR, a technique that recovers A3-edited DNA at lower PCR denaturation temperatures (Td) [64]. As can be seen in Fig 4A and 4B, A3DE expression shifted to lower temperatures the last Td at which DNA could be recovered in a dose dependent manner – samples highlighted by an asterisk. As anticipated from the GST pulldown and immunofluorescence experiments (Fig 1 and 2), co-transfection of A3DE did not impact HBV DNA editing by A3B (Fig 4C). However, A3DE did antagonize A3BnA editing of HBV DNA editing (Fig 3D). Those samples highlighted by an asterisk were cloned and sequenced and bore the hallmarks of A3 editing, i.e. a preference for editing at 5'TpC for A3B, A3BnA and A3F and 5'CpC for A3G (data not shown).

APOBEC3DE enhances HBV replication

To assess the relative expression levels of *A3* genes in Huh7, a TaqMan transcriptional study was made in presence or absence of HBV (Fig 5A). Of the seven *A3* genes analysed, using *RPL13A* as reference, Huh7 cells showed significantly higher *A3B-A3G* and lower *A3A* and *A3H* expression. Interestingly, the relative *A3* genes expression was similar in presence and absence of HBV, suggesting that viral proteins do not induce *A3* genes expression. Moreover,

we observed an identical expression of *A3DE*, *A3F* and *A3G*, showing that endogenous level of A3DE was not sufficient to counteract A3G or A3F levels.

In order to quantitate the impact of A3DE expression on HBV levels, Huh7 cells were co-transfected by pCayw with increasing amounts of A3DE. As can be seen on Fig 5B, 0.25 µg and 0.75 µg of A3DE increased HBV DNA load in culture supernatants, presumably by counteracting endogenous levels of A3 enzymes. When A3F or A3G was co-transfected, HBV DNA levels dropped by factors of 5-7 (Fig 5B). Increasing amounts of A3DE countered well restriction by A3F and to a lesser extent A3G, (Fig 5B), in keeping with the amino acid homology between A3D vs. A3F N-terminal domain (67%) and A3D vs. A3G N-terminal domain (44%). Inversely, the presence of A3DE siRNA knockdown, when compared to siRNA control in Huh7 cells, restricts HBV DNA by factors of ~20 by disrupting endogenous A3DE transcripts (Fig 5C). Efficiency of A3DE siRNA was performed and presented in Fig 5D. Despite antagonizing A3F and A3G restriction of HBV DNA synthesis, A3DE heterodimer formation had no effect on steady state A3F and A3G protein levels (Fig. 5E) meaning that A3DE restriction follows from dilution of the number of active sites in multimers and not by degradation of A3F or A3G.

APOBEC3DEC320Y hyperedits HBV DNA

Early reports suggested that the A3DEC320 protein could restrict human immunodeficiency type 1 replication and even edit ssDNA [57, 65, 66]. No hypermutated viral sequences were given to support the contention, although editing frequencies of the order of ~0.02% were reported [57], more than ~1000 fold lower than for classical hypermutants. To explore this, we used very sensitive 3DPCR [64], which is able to pick up a highly A3-edited molecule among up to 10⁴ unedited molecules. Transfection of QT6 cells with pCayw and an

A3DEC320–V5 expression construct failed to recover hyperedited HBV X gene DNA. By contrast, for an A3DEC320Y construct, 3DPCR products corresponding were recovered as low as 84.8°C for the A3DEC320Y compared to 89.6°C for the A3DE or the empty plasmid control (Fig 6A). 3DPCR products from 88.1°C and 86.5°C reactions were cloned and sequenced. Extensive and monotonously C->T edited HBV sequences were recovered, some with up to 90% of cytidine targets edited (Fig 6B, D). The dinucleotide context was uniform with a strong preference for 5'TpC, typical of A3A, A3B, A3C and A3H enzymes (Fig 6C, [30, 36, 37]) followed by GpC, which is unusual for A3 enzymes. Interestingly A3F editing of the same region of the HBV genome follows the same pattern (TpC>GpC>CpC>ApC), which is consistent with their phylogenic relatedness [40]. Interestingly, A3DEC320Y could also edit cytoplasmic mitochondrial DNA with a similar dinucleotide context to shown for HBV DNA, showing that A3DEC320Y doesn't need to be encapsidated for ssDNA editing (data not shown).

The above findings show that A3DEC320 does not edit HBV DNA. A non-editing restriction phenomenon along the lines of some catalytically inactive A3G mutants cannot be excluded [39, 53-55]. However, in the context of HBV, this is unlikely to be strong for A3DE over-expression led to higher HBV DNA levels (Fig 5A). Alternatively, any non-editing restriction effect is less pronounced than antagonism of the A3F and A3G restriction factors.

The data in Fig 6 show that A3DEC320Y was indeed capable of hyperediting HBV DNA (Fig 6C). Intrigued by this and the realization that A3DEC320 activity appeared to be pro-viral and facilitate HBV replication, we genotyped 41 cases of liver cirrhosis from a previous study [12] and a further 60 samples from HBV associated hepatocellular carcinoma as well as a variety of liver derived cell lines including hepatocellular carcinoma cell lines (S1 Table). We

found A3DEC320 in all samples analyzed. The C320Y substitution corresponds to the G>A SNP rs369944754. Screening the 1000 genomes database showed that the A3DEC320 allele was exclusively present at a very low frequency ~0.1% (www.1000genomes.org) indicating that A3DE is not polymorphic at this site.

It is ironic that A3DE, a member of a cluster of viral restriction factors, actually enhances HBV replication, the more so as it is up-regulated in alcoholic and virally-associated liver cirrhosis [12]. The finding is reminiscent of some interferon stimulated genes (ISGs) which, when analyzed individually, can enhance viral replication. For example, ADAR1, which deaminates adenosine residues in viral dsRNA, can enhance HIV-1 and Chikungunya virus replication by 50-75% [67]. Earlier work showed ADAR1 to be essential to the life cycle of hepatitis D virus [68]. Closer to home, A3A cytidine deaminase could enhance West Nile and Venezuelan equine encephalitis viruses by 50-60%, although the mechanism is unknown [67]. While not fully titrated out, A3DE enhancement of HBV DNA in Fig 5A is ~70%. The A3DEC320Y polymorphism seems to be fully fixed in the human population suggesting that there are evolutionary advantages to being catalytically inactive. Alternatively, it could have been fixed via the strong bottleneck that human precursor populations went through [69]. It is a recent experiment in evolution for only the gorilla harbors a similar substitution (C320). All other monkeys and apes encode Y320 with the macaque A3DE protein being demonstrably catalytic [60]. That it occurs in another great ape lineages argues for an evolutionary advantage for the C320 allele under some circumstances.

With cross-talk between A3DE, A3F and A3G, the role of some of these innate restriction factors is clearly more subtle than previously thought. The A3DE proximal promoter is closest to that of A3C (53% identity over the proximal 700 bp) neither of which are interferon

sensitive so marking it out from A3A and A3G, which are clearly ISGs [21-25]. Possibly
interferon induction of A3G is a means of overcoming basal levels of A3DE. Like the other
A3 genes, A3DE is particularly well expressed in hematopoietic cells (www.biogps.org). It is
possible that it lowers the thresholds levels of A3F and A3G so facilitating the function of the
HIV Vif protein.

Materials and Methods

Patients

The patients with cirrhosis have been already described [12]. All samples, chosen in
retrospective collections of patients affected with HCCs have been previously published [70-
72]. The recruitment of patients was respecting the declaration of Helsinki and was approved
by an Institutional Review Board (RBM n°2005-019). Informed consent was obtained for
each patient.

Cells lines and culture.

Huh7, HeLa and HEK293T cells were maintained in DMEM medium supplemented with
10% FCS, 100 units/ml penicillin, and 100 µg/ml streptomycin (GIBCO/BRL) at 37°C in 5%
CO₂. QT6 cells were maintained in Ham's medium supplemented with 100 U/ml penicillin
and 100 µg/ml streptomycin (GIBCO), 2 mM glutamine, 5% tryptose phosphate, 1% chicken
serum and 10% FCS at 37°C in 5% CO₂.

Approximately 450,000 QT6 cells were seeded in 6 well plates and transfected 24 h later by 2
µg of total plasmid using JetPrime (Polyplus Transfection TM). After 48 h, total DNA was
extracted using EPICENTRE kit. For imaging, 50,000 HeLa cells were seeded in Lab-Tek

and transfected 24 h later by 1 µg of total plasmid using FuGENE HD Transfection Reagent (Roche Applied Science). For qPCR, 450,000 Huh7 cells were seeded in 6 wells plates and transfected 24 h later by 2 µg of total plasmid using JetPrime (Polyplus Transfection TM). Twenty-four hours post transfection, the medium was changed. After 48h or 7 days of culture, supernatant was treated with 10 U of DNase I and incubated 2h at 37°C and then heat inactivated. DNA was extracted from the supernatant using EPICENTRE kit (Illumina).

Plasmids

cDNA encodings for APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3DE, APOBEC3G, APOBEC3F, APOBEC3G and APOBEC3H were amplified by standard PCR (*Taq* Platinum; Invitrogen) and cloned in pcDNA3.1/V5-His TOPO vector using pcDNA 3.1/V5-His TOPO Expression Kit (Invitrogen), and by in vitro recombination in pDONR207 using Gateway technology (Invitrogen). PCR primers displayed 20–30 specific nucleotides matching ORF extremities so that their T_m is close to 60 °C. The 5' ends of Gateway forward primers were fused to attB1.1 recombination sequence 5'GGGGACAACCTTTGTACAAAAAAGTTGGCATG-3', whereas reverse primers were fused to attB2.1 recombination sequence 5'-GGGGACAACCTTTGTACAAGAAAGTTGGTTA-3'. Recombination of PCR products into pDONR207 was performed following the manufacturer's recommendation (BP cloning reaction; Invitrogen). All constructs were transformed and amplified in Escherichia coli TOP10 strain. APOBEC3BnA was synthesized retaining intron 7 and cloned in pDONR207 like the others. A3DE siRNA is a mixture of 5'GUUUGAGAACCACGCAGAA and 5'UUCUGCGUGGUUCUCAAAC and siRNA universal negative control (SIC001) were from Sigma-Aldrich.

Co-affinity Purification and protein extraction

To perform co-affinity purification experiments, cloned ORFs were transfected from pDONR207 to perform co-affinity purification experiments; cloned ORFs were transferred from pDONR207 to pDEST27 expression vector (Invitrogen) to achieve GST fusion and to pCI-neo-3xFLAG vector [21] for 3xFLAG fusion. Tagged proteins were expressed in human HEK-293T. Briefly, 5 10^5 HEK-293T cells were seeded in 6-well plates and transfected 24 h later with 300 ng of the GST constructs and 100 ng of the 3xFLAG constructs. For co-affinity purification and protein extraction, two days after transfection, HEK-293T cells were washed in phosphate-buffered saline (PBS) and then resuspended in lysis buffer (0.5% Nonidet P-40, 20 mM Tris-HCl, pH 7.4, 120 mM NaCl, and 1 mM EDTA) supplemented with Complete Protease Inhibitor Mixture (Roche Applied Science). Cell lysates were incubated on ice for 20 min and then clarified by centrifugation at 14,000 g for 30 min. For pulldown analysis, 400 μ g of protein extracts were incubated for 1h at 4°C with 25 μ l of glutathione-Sepharose beads (Amersham Biosciences) to purify GST-tagged proteins. Beads were then washed three times in ice-cold lysis buffer, and proteins were recovered by boiling in denaturing loading buffer (Invitrogen).

Western Blotting

Purified complexes and protein extracts were resolved by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) on 4–12% NuPAGE Bis-Tris gels with MOPS running buffer (Invitrogen), and transferred to a nitrocellulose membrane. 3xFLAG, V5 and GST-tagged proteins were detected using standard immunoblotting techniques. Membranes were blotted with a mouse monoclonal HRP conjugated anti-3xFLAG antibody (M2; Sigma) or a rabbit

polyclonal anti-GST antibody (Sigma) or a mouse monoclonal HRP conjugated anti-V5 antibody (Invitrogen). Secondary anti- rabbit HRP-conjugated antibody was from Amersham Biosciences.

DNA extraction, PCR and 3D-PCR

Total DNA from transfected cells was extracted using the MasterPure™ complete DNA and RNA purification kit (Epicentre) and resuspended in 30 µl sterile water. Non-tumor tissue samples were submitted to digestion in SDS/proteinase K buffer at 37°C for 6 to 12h, followed by two phenol and one chloroform extractions. DNA was then ethanol-precipitated and resuspended in TE buffer. For the amplification of HBV X gene, the first round reaction parameters were 95°C for 5 min followed by 35 cycles (95°C for 1 min, 50°C for 30 sec, 72°C for 5 min) and finally 10 min at 72°C with primers 5'CGCAAATATACATCGTATCCAT and 5'AAGAGTYTYTTATGTAAGAGCYT. For the second round 3DPCR were performed using the equivalent of 0.5 µg of first round reaction as input using Eppendorf gradient Mastercycler S. The reaction parameters were 80-92°C for 5 min followed by 35 cycles (80-92°C for 1 min, 45°C for 30 sec and 72°C for 2 min) and finally 10 min at 72°C. and primers were used as forward and reverse primers respectively 5'ATGGCTGCTARGCTGTGCTGCCAA and 5'AAGTGCACACGGTYYGCGAGAT, where Y is T/C and R is A/G. All amplifications were performed using first round standard PCR followed by nested 3D-PCR. PCR was performed with 2.5 units of *Taq* (Eurobio) DNA polymerase per reaction.

For the amplification of exon 6 of A3DE, 25ng of genomic DNA were subjected to 35 cycles of amplification according to a "step-down" protocol in a Techne Genius thermal cycler (OSI, Saint-Quentin-en-Yvelines, France). A first round of amplification with primers anchored in

APOBEC3D-specific intron sequence was used followed by a second internal PCR. First Amplification was performed with primers 3DEes, 5'ACCCCAGCTTGGGCAACAGG and 3DEer, 5'TGCAGCTGGCGCCGGCCCCA. Reactions were performed in 25µl of a mixture containing 1X PCR buffer (50 mM KCl, 10 mM Tris pH 8.9, 0.1% Tween 20, 1.5 mM MgCl₂), 200 µM dNTP, 20 pmoles of each primer, 1 unit of Taq polymerase. Conditions were 95°C for 5 min followed by 35 cycles (95°C for 30 sec, 52°C for 30 sec, 72°C for 1 min) and finally 10 min. Nested PCR amplification was performed with primers 3DE320p, 5'TGTCTCCTAACACAAACTACG and 3DE320j, 5'CCCCTCCTGGTAATCTGTATCCCAGAACTAG. PCR products were purified from agarose gels (Qiaex II kit, Qiagen, France) and ligated into the TOPO TA cloning vector (Invitrogen, France). Sequencing was outsourced to GATC. All mutations were verified on the chromatogram.

Quantitative PCR

HBV DNA levels were quantified by qPCR using an Eppendorf Realplex². Reactions were performed in 96-well plates with each containing 12.5 µL 2x probe Master Mix (Roche), 2 µl primers (5 µm each), 0.33 µL UPL probe (Roche), 5.17 µL H₂O and 5 µl cDNA. The reaction parameters were 95°C 10min followed by 45 cycles (95°C 15 sec, 58°C 15 sec, 68°C 30 sec). 5' AAGAGTYITYTTATGTAAGAGCYT and 5' CGCAAATATACATCGTATCCAT primers were used as forward and reverse primers respectively for HBV DNA quantification, in combination with UPL125 probe.

Confocal microscopy

After PBS washing, cell were fixed in paraformaldehyde 4% for 20 min at room temperature. Then, cells were fixed in 50/50 methanol/acetone for 20 min at room temperature. The mouse

anti-V5 antibody (Invitrogen) and rabbit anti Flag-antibody (Sigma) were incubated at 1/200 for 1h at room temperature followed by incubation with fluoerescein –conjugated secondary antibody (Alexa 488 and Alexa 555, Invitrogen) for 1h at room temperature. Slides were mounted with Vectashield-DAPI. Confocal Imaging was performed using a Leica-TCS-SP5. X1.2 WATER optics was used.

Colocalization analyse was performed with Huygens Essential for mac version 4.3.1p3 (Scientific Volume Imaging B.V.). Images were first deconvoluted. Then, analysis was performed using the Costes method.

Acknowledgements

This work was supported by grants from the Institut Pasteur, ANRS and CNRS. MSB was supported by a bursary from la Ligue Nationale contre le Cancer. We thank Yong-Hui Zheng for providing APOBEC3DE plasmid.

374 **References**

- 375 1. Jarmuz A, Chester A, Bayliss J, Gisbourne J, Dunham I, Scott J, et al. An anthropoid-
376 specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics*.
377 2002;79(3):285-96. PubMed PMID: 11863358.
- 378 2. Conticello SG, Thomas CJ, Petersen-Mahrt SK, Neuberger MS. Evolution of the
379 AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol Biol Evol*.
380 2005;22(2):367-77.
- 381 3. Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, et al. DNA
382 deamination mediates innate immunity to retroviral infection. *Cell*. 2003;113(6):803-9.
383 PubMed PMID: 12809610.
- 384 4. Lecossier D, Bouchonnet F, Clavel F, Hance AJ. Hypermutation of HIV-1 DNA in the
385 absence of the Vif protein. *Science*. 2003;300(5622):1112. PubMed PMID: 12750511.
- 386 5. Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D. Broad antiretroviral defence
387 by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature*.
388 2003;424(6944):99-103. PubMed PMID: 12808466.
- 389 6. Mariani R, Chen D, Schrofelbauer B, Navarro F, König R, Bollman B, et al. Species-
390 specific exclusion of APOBEC3G from HIV-1 virions by Vif. *Cell*. 2003;114(1):21-31.
391 PubMed PMID: 12859895.
- 392 7. Sheehy AM, Gaddis NC, Choi JD, Malim MH. Isolation of a human gene that inhibits
393 HIV-1 infection and is suppressed by the viral Vif protein. *Nature*. 2002;418(6898):646-
394 50. PubMed PMID: 12167863.
- 395 8. Lucifora J, Xia Y, Reisinger F, Zhang K, Stadler D, Cheng X, et al. Specific and
396 nonhepatotoxic degradation of nuclear hepatitis B virus cccDNA. *Science*.
397 2014;343(6176):1221-8. Epub 2014/02/22. doi: 10.1126/science.1243462. PubMed PMID:
398 24557838.
- 399 9. Noguchi C, Ishino H, Tsuge M, Fujimoto Y, Imamura M, Takahashi S, et al. G to A
400 hypermutation of hepatitis B virus. *Hepatology*. 2005;41(3):626-33. Epub 2005/02/24. doi:
401 10.1002/hep.20580. PubMed PMID: 15726649.
- 402 10. Suspène R, Guétard D, Henry M, Sommer P, Wain-Hobson S, Vartanian JP.
403 Extensive editing of both hepatitis B virus DNA strands by APOBEC3 cytidine deaminases
404 in vitro and in vivo. *Proc Natl Acad Sci USA*. 2005;102(23):8321-6. PubMed PMID:
405 15919829.
- 406 11. Turelli P, Mangeat B, Jost S, Vianin S, Trono D. Inhibition of hepatitis B virus replication
407 by APOBEC3G. *Science*. 2004;303(5665):1829. PubMed PMID: 15031497.
- 408 12. Vartanian JP, Henry M, Marchio A, Suspène R, Aynaud MM, Guétard D, et al. Massive
409 APOBEC3 editing of hepatitis B viral DNA in cirrhosis. *PLoS Pathog*.
410

- 2010;6(5):e1000928. Epub 2010/06/05. doi: 10.1371/journal.ppat.1000928. PubMed PMID: 20523896; PubMed Central PMCID: PMC2877740.
13. Suspène R, Aynaud MM, Koch S, Pasdeloup D, Labetoulle M, Gaertner B, et al. Genetic editing of herpes simplex virus 1 and Epstein-Barr herpesvirus genomes by human APOBEC3 cytidine deaminases in culture and in vivo. *J Virol.* 2011;85(15):7594-602. doi: JVI.00290-11 [pii] 10.1128/JVI.00290-11. PubMed PMID: 21632763; PubMed Central PMCID: PMC3147940.
 14. Vartanian JP, Guétard D, Henry M, Wain-Hobson S. Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. *Science.* 2008;320(5873):230-3. Epub 2008/04/12. doi: 320/5873/230 [pii] 10.1126/science.1153201. PubMed PMID: 18403710.
 15. Wang Z, Wakae K, Kitamura K, Aoyama S, Liu G, Koura M, et al. APOBEC3 deaminases induce hypermutation in human papillomavirus 16 DNA upon beta interferon stimulation. *J Virol.* 2014;88(2):1308-17. Epub 2013/11/15. doi: 10.1128/JVI.03091-13. PubMed PMID: 24227842; PubMed Central PMCID: PMC3911654.
 16. Warren CJ, Xu T, Guo K, Griffin LM, Westrich JA, Lee D, et al. APOBEC3A functions as a restriction factor of human papillomavirus. *J Virol.* 2015;89(1):688-702. Epub 2014/10/31. doi: 10.1128/JVI.02383-14. PubMed PMID: 25355878; PubMed Central PMCID: PMC4301161.
 17. Bogerd HP, Wiegand HL, Hulme AE, Garcia-Perez JL, O'Shea KS, Moran JV, et al. Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc Natl Acad Sci USA.* 2006;103(23):8780-5. Epub 2006/05/27. doi: 0603313103 [pii] 10.1073/pnas.0603313103. PubMed PMID: 16728505.
 18. Chen H, Lilley CE, Yu Q, Lee DV, Chou J, Narvaiza I, et al. APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. *Curr Biol.* 2006;16(5):480-5. Epub 2006/03/11. doi: S0960-9822(06)01040-2 [pii] 10.1016/j.cub.2006.01.031. PubMed PMID: 16527742.
 19. Muckenfuss H, Hamdorf M, Held U, Perkovic M, Lower J, Cichutek K, et al. APOBEC3 proteins inhibit human LINE-1 retrotransposition. *J Biol Chem.* 2006;281(31):22161-72. Epub 2006/06/01. doi: M601716200 [pii] 10.1074/jbc.M601716200. PubMed PMID: 16735504.
 20. Suspène R, Sommer P, Henry M, Ferris S, Guétard D, Pochet S, et al. APOBEC3G is a single-stranded DNA cytidine deaminase and functions independently of HIV reverse transcriptase. *Nucleic Acids Res.* 2004;32(8):2421-9.
 21. Aynaud MM, Suspène R, Vidalain PO, Mussil B, Guétard D, Tangy F, et al. Human Tribbles 3 protects nuclear DNA from cytidine deamination by APOBEC3A. *J Biol Chem.* 2012;287(46):39182-92. Epub 2012/09/15. doi: 10.1074/jbc.M112.372722. PubMed PMID: 22977230.
 22. Bonvin M, Achermann F, Greeve I, Stroka D, Keogh A, Inderbitzin D, et al. Interferon-inducible expression of APOBEC3 editing enzymes in human hepatocytes and inhibition

- of hepatitis B virus replication. *Hepatology*. 2006;43(6):1364-74. PubMed PMID: 16729314.
23. Koning FA, Newman EN, Kim EY, Kunstman KJ, Wolinsky SM, Malim MH. Defining APOBEC3 expression patterns in human tissues and hematopoietic cell subsets. *J Virol*. 2009;83(18):9474-85. Epub 2009/07/10. doi: JVI.01089-09 [pii] 10.1128/JVI.01089-09. PubMed PMID: 19587057.
 24. Peng G, Lei KJ, Jin W, Greenwell-Wild T, Wahl SM. Induction of APOBEC3 family proteins, a defensive maneuver underlying interferon-induced anti-HIV-1 activity. *J Exp Med*. 2006;203(1):41-6. PubMed PMID: 16418394.
 25. Stenglein MD, Burns MB, Li M, Lengyel J, Harris RS. APOBEC3 proteins mediate the clearance of foreign DNA from human cells. *Nat Struct Mol Biol*. 2010;17(2):222-9. Epub 2010/01/12. doi: nsmb.1744 [pii] 10.1038/nsmb.1744. PubMed PMID: 20062055.
 26. Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, Leonard B, et al. APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature*. 2013;494(7437):366-70. Epub 2013/02/08. doi: 10.1038/nature11881. PubMed PMID: 23389445.
 27. Caval V, Suspène R, Shapira M, Vartanian JP, Wain-Hobson S. A prevalent cancer susceptibility APOBEC3A hybrid allele bearing APOBEC3B 3'UTR enhances chromosomal DNA damage. *Nature communications*. 2014;5:5129. Epub 2014/10/10. doi: 10.1038/ncomms6129. PubMed PMID: 25298230.
 28. Mussil B, Suspène R, Aynaud MM, Gaurvit A, Vartanian JP, Wain-Hobson S. Human APOBEC3A isoforms translocate to the nucleus and induce DNA double strand breaks leading to cell stress and death. *PLoS ONE*. 2013;8(8):e73641. Epub 2013/08/27. doi: 10.1371/journal.pone.0073641. PubMed PMID: 23977391; PubMed Central PMCID: PMC3748023.
 29. Shinohara M, Io K, Shindo K, Matsui M, Sakamoto T, Tada K, et al. APOBEC3B can impair genomic stability by inducing base substitutions in genomic DNA in human cells. *Scientific reports*. 2012;2:806. Epub 2012/11/15. doi: 10.1038/srep00806. PubMed PMID: 23150777; PubMed Central PMCID: PMC3496164.
 30. Suspène R, Aynaud M, Guétard D, Henry M, Eckhoff G, Marchio A, et al. Somatic hypermutation of human mitochondrial and nuclear DNA by APOBEC3 cytidine deaminases, a pathway for DNA catabolism. *Proc Natl Acad Sci USA*. 2011;108(12):4858-63.
 31. Carpenter MA, Li M, Rathore A, Lackey L, Law EK, Land AM, et al. Methylcytosine and Normal Cytosine Deamination by the Foreign DNA Restriction Enzyme APOBEC3A. *J Biol Chem*. 2012;287(41):34801-8. Epub 2012/08/17. doi: 10.1074/jbc.M112.385161. PubMed PMID: 22896697; PubMed Central PMCID: PMC3464582.
 32. Suspène R, Aynaud MM, Vartanian JP, Wain-Hobson S. Efficient deamination of 5-methylcytidine and 5-substituted cytidine residues in DNA by human APOBEC3A cytidine deaminase. *PLoS ONE*. 2013;8(6):e63461. Epub 2013/07/11. doi:

- 10.1371/journal.pone.0063461. PubMed PMID: 23840298; PubMed Central PMCID: PMC3688686.
33. Wijesinghe P, Bhagwat AS. Efficient deamination of 5-methylcytosines in DNA by human APOBEC3A, but not by AID or APOBEC3G. *Nucleic Acids Res.* 2012;40(18):9206-17. Epub 2012/07/17. doi: 10.1093/nar/gks685. PubMed PMID: 22798497; PubMed Central PMCID: PMC3467078.
 34. Caval V, Suspene R, Vartanian JP, Wain-Hobson S. Orthologous mammalian APOBEC3A cytidine deaminases hypermutate nuclear DNA. *Mol Biol Evol.* 2014;31(2):330-40. Epub 2013/10/29. doi: 10.1093/molbev/mst195. PubMed PMID: 24162735.
 35. Landry S, Narvaiza I, Linfesty DC, Weitzman MD. APOBEC3A can activate the DNA damage response and cause cell-cycle arrest. *EMBO reports.* 2011;12(5):444-50. Epub 2011/04/05. doi: 10.1038/embor.2011.46. PubMed PMID: 21460793; PubMed Central PMCID: PMC3090015.
 36. Bishop KN, Holmes RK, Sheehy AM, Davidson NO, Cho SJ, Malim MH. Cytidine deamination of retroviral DNA by diverse APOBEC proteins. *Curr Biol.* 2004;14(15):1392-6. PubMed PMID: 15296758.
 37. Beale RC, Petersen-Mahrt SK, Watt IN, Harris RS, Rada C, Neuberger MS. Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo. *J Mol Biol.* 2004;337(3):585-96. PubMed PMID: 15019779.
 38. Hache G, Liddament MT, Harris RS. The retroviral hypermutation specificity of APOBEC3F and APOBEC3G is governed by the C-terminal DNA cytosine deaminase domain. *J Biol Chem.* 2005;280(12):10920-4. Epub 2005/01/14. doi: 10.1074/jbc.M500382200. PubMed PMID: 15647250.
 39. Newman EN, Holmes RK, Craig HM, Klein KC, Lingappa JR, Malim MH, et al. Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. *Curr Biol.* 2005;15(2):166-70. Epub 2005/01/26. doi: S0960982204010498 [pii] 10.1016/j.cub.2004.12.068. PubMed PMID: 15668174.
 40. Henry M, Guetard D, Suspene R, Rusniok C, Wain-Hobson S, Vartanian JP. Genetic editing of HBV DNA by monodomain human APOBEC3 cytidine deaminases and the recombinant nature of APOBEC3G. *PLoS One.* 2009;4(1):e4277. Epub 2009/01/27. doi: 10.1371/journal.pone.0004277. PubMed PMID: 19169351.
 41. Bohn MF, Shandilya SM, Silvas TV, Nalivaika EA, Kouno T, Kelch BA, et al. The ssDNA Mutator APOBEC3A Is Regulated by Cooperative Dimerization. *Structure.* 2015;23(5):903-11. Epub 2015/04/29. doi: 10.1016/j.str.2015.03.016. PubMed PMID: 25914058.
 42. Friew YN, Boyko V, Hu WS, Pathak VK. Intracellular interactions between APOBEC3G, RNA, and HIV-1 Gag: APOBEC3G multimerization is dependent on its association with RNA. *Retrovirology.* 2009;6:56. Epub 2009/06/06. doi: 10.1186/1742-4690-6-56. PubMed PMID: 19497112; PubMed Central PMCID: PMC2700067.

43. Huthoff H, Autore F, Gallois-Montbrun S, Fraternali F, Malim MH. RNA-dependent oligomerization of APOBEC3G is required for restriction of HIV-1. *PLoS Pathog.* 2009;5(3):e1000330. Epub 2009/03/07. doi: 10.1371/journal.ppat.1000330. PubMed PMID: 19266078; PubMed Central PMCID: PMC2646141.
44. Koyama T, Arias JF, Iwabu Y, Yokoyama M, Fujita H, Sato H, et al. APOBEC3G oligomerization is associated with the inhibition of both Alu and LINE-1 retrotransposition. *PLoS ONE.* 2013;8(12):e84228. Epub 2013/12/25. doi: 10.1371/journal.pone.0084228. PubMed PMID: 24367644; PubMed Central PMCID: PMC3868573.
45. Stauch B, Hofmann H, Perkovic M, Weisel M, Kopietz F, Cichutek K, et al. Model structure of APOBEC3C reveals a binding pocket modulating ribonucleic acid interaction required for encapsidation. *Proc Natl Acad Sci U S A.* 2009;106(29):12079-84. Epub 2009/07/08. doi: 10.1073/pnas.0900979106. PubMed PMID: 19581596; PubMed Central PMCID: PMC2715496.
46. Wiegand HL, Doehle BP, Bogerd HP, Cullen BR. A second human antiretroviral factor, APOBEC3F, is suppressed by the HIV-1 and HIV-2 Vif proteins. *EMBO J.* 2004;23(12):2451-8. PubMed PMID: 15152192.
47. Brar SS, Sacho EJ, Tessmer I, Croteau DL, Erie DA, Diaz M. Activation-induced deaminase, AID, is catalytically active as a monomer on single-stranded DNA. *DNA repair.* 2008;7(1):77-87. Epub 2007/09/25. doi: 10.1016/j.dnarep.2007.08.002. PubMed PMID: 17889624; PubMed Central PMCID: PMC2693009.
48. Salter JD, Krucinska J, Raina J, Smith HC, Wedekind JE. A hydrodynamic analysis of APOBEC3G reveals a monomer-dimer-tetramer self-association that has implications for anti-HIV function. *Biochemistry.* 2009;48(45):10685-7. Epub 2009/10/21. doi: 10.1021/bi901642c. PubMed PMID: 19839647.
49. Chiu YL, Witkowska HE, Hall SC, Santiago M, Soros VB, Esnault C, et al. High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition. *Proc Natl Acad Sci U S A.* 2006;103(42):15588-93. Epub 2006/10/13. doi: 0604524103 [pii] 10.1073/pnas.0604524103. PubMed PMID: 17030807.
50. McDougall WM, Okany C, Smith HC. Deaminase activity on single-stranded DNA (ssDNA) occurs in vitro when APOBEC3G cytidine deaminase forms homotetramers and higher-order complexes. *J Biol Chem.* 2011;286(35):30655-61. Epub 2011/07/09. doi: 10.1074/jbc.M111.269506. PubMed PMID: 21737457; PubMed Central PMCID: PMC3162426.
51. Wedekind JE, Gillilan R, Janda A, Krucinska J, Salter JD, Bennett RP, et al. Nanostructures of APOBEC3G support a hierarchical assembly model of high molecular mass ribonucleoprotein particles from dimeric subunits. *J Biol Chem.* 2006;281(50):38122-6. Epub 2006/11/03. doi: 10.1074/jbc.C600253200. PubMed PMID: 17079235; PubMed Central PMCID: PMC1847398.

52. Dang Y, Abudu A, Son S, Harjes E, Spearman P, Matsuo H, et al. Identification of a single amino acid required for APOBEC3 antiretroviral cytidine deaminase activity. *J Virol.* 2011;85(11):5691-5. Epub 2011/03/25. doi: 10.1128/JVI.00243-11. PubMed PMID: 21430060; PubMed Central PMCID: PMC3094948.
53. Holmes RK, Koning FA, Bishop KN, Malim MH. APOBEC3F can inhibit the accumulation of HIV-1 reverse transcription products in the absence of hypermutation. Comparisons with APOBEC3G. *J Biol Chem.* 2007;282(4):2587-95. Epub 2006/11/24. doi: 10.1074/jbc.M607298200. PubMed PMID: 17121840.
54. Nguyen DH, Gummuluru S, Hu J. Deamination-independent inhibition of hepatitis B virus reverse transcription by APOBEC3G. *J Virol.* 2007;81(9):4465-72. Epub 2007/02/23. doi: JVI.02510-06 [pii] 10.1128/JVI.02510-06. PubMed PMID: 17314171.
55. Rosler C, Kock J, Kann M, Malim MH, Blum HE, Baumert TF, et al. APOBEC-mediated interference with hepadnavirus production. *Hepatology.* 2005;42(2):301-9. Epub 2005/07/19. doi: 10.1002/hep.20801. PubMed PMID: 16025511.
56. LaRue RS, Jonsson SR, Silverstein KA, Lajoie M, Bertrand D, El-Mabrouk N, et al. The artiodactyl APOBEC3 innate immune repertoire shows evidence for a multi-functional domain organization that existed in the ancestor of placental mammals. *BMC Mol Biol.* 2008;9:104. Epub 2008/11/20. doi: 10.1186/1471-2199-9-104. PubMed PMID: 19017397; PubMed Central PMCID: PMC2612020.
57. Dang Y, Wang X, Esselman WJ, Zheng YH. Identification of APOBEC3DE as another antiretroviral factor from the human APOBEC family. *J Virol.* 2006;80(21):10522-33. PubMed PMID: 16920826.
58. Peng G, Greenwell-Wild T, Nares S, Jin W, Lei KJ, Rangel ZG, et al. Myeloid differentiation and susceptibility to HIV-1 are linked to APOBEC3 expression. *Blood.* 2007;110(1):393-400. Epub 2007/03/21. doi: 10.1182/blood-2006-10-051763. PubMed PMID: 17371941; PubMed Central PMCID: PMC1896122.
59. Chaipan C, Smith JL, Hu WS, Pathak VK. APOBEC3G restricts HIV-1 to a greater extent than APOBEC3F and APOBEC3DE in human primary CD4+ T cells and macrophages. *J Virol.* 2013;87(1):444-53. Epub 2012/10/26. doi: 10.1128/JVI.00676-12. PubMed PMID: 23097438; PubMed Central PMCID: PMC3536366.
60. Duggal NK, Malik HS, Emerman M. The breadth of antiviral activity of Apobec3DE in chimpanzees has been driven by positive selection. *J Virol.* 2011;85(21):11361-71. Epub 2011/08/13. doi: 10.1128/JVI.05046-11. PubMed PMID: 21835794; PubMed Central PMCID: PMC3194980.
61. Bennett RP, Salter JD, Liu X, Wedekind JE, Smith HC. APOBEC3G subunits self-associate via the C-terminal deaminase domain. *J Biol Chem.* 2008;283(48):33329-36. Epub 2008/10/10. doi: 10.1074/jbc.M803726200. PubMed PMID: 18842592; PubMed Central PMCID: PMC2586250.

62. Chelico L, Prochnow C, Erie DA, Chen XS, Goodman MF. Structural model for deoxycytidine deamination mechanisms of the HIV-1 inactivation enzyme APOBEC3G. *J Biol Chem.* 2010;285(21):16195-205. Epub 2010/03/10. doi: M110.107987 [pii] 10.1074/jbc.M110.107987. PubMed PMID: 20212048.
63. Shlyakhtenko LS, Lushnikov AY, Li M, Lackey L, Harris RS, Lyubchenko YL. Atomic force microscopy studies provide direct evidence for dimerization of the HIV restriction factor APOBEC3G. *J Biol Chem.* 2011;286(5):3387-95. Epub 2010/12/03. doi: 10.1074/jbc.M110.195685. PubMed PMID: 21123176; PubMed Central PMCID: PMC3030345.
64. Suspène R, Henry M, Guillot S, Wain-Hobson S, Vartanian JP. Recovery of APOBEC3-edited human immunodeficiency virus G->A hypermutants by differential DNA denaturation PCR. *J Gen Virol.* 2005;86(Pt 1):125-9.
65. Hultquist JF, Lengyel JA, Refsland EW, Larue RS, Lackey L, Brown WL, et al. Human and Rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H Demonstrate a Conserved Capacity to Restrict Vif-deficient HIV-1. *J Virol.* 2011;84(19):11220-112234. Epub 2011/08/13. doi: JVI.05238-11 [pii] 10.1128/JVI.05238-11. PubMed PMID: 21835787.
66. Sato K, Takeuchi JS, Misawa N, Izumi T, Kobayashi T, Kimura Y, et al. APOBEC3D and APOBEC3F potentially promote HIV-1 diversification and evolution in humanized mouse model. *PLoS Pathog.* 2014;10(10):e1004453. Epub 2014/10/21. doi: 10.1371/journal.ppat.1004453. PubMed PMID: 25330146; PubMed Central PMCID: PMC4199767.
67. Schoggins JW, Wilson SJ, Panis M, Murphy MY, Jones CT, Bieniasz P, et al. A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature.* 2011;472(7344):481-5. Epub 2011/04/12. doi: 10.1038/nature09907. PubMed PMID: 21478870; PubMed Central PMCID: PMC3409588.
68. Polson AG, Bass BL, Casey JL. RNA editing of hepatitis delta virus antigenome by dsRNA-adenosine deaminase. *Nature.* 1996;380(6573):454-6. Epub 1996/04/04. doi: 10.1038/380454a0. PubMed PMID: 8602246.
69. Gagneux P, Wills C, Gerloff U, Tautz D, Morin PA, Boesch C, et al. Mitochondrial sequences show diverse evolutionary histories of African hominoids. *Proceedings of the National Academy of Sciences USA.* 1999;96(9):5077-82.
70. Pineau P, Marchio A, Battiston C, Cordina E, Russo A, Terris B, et al. Chromosome instability in human hepatocellular carcinoma depends on p53 status and aflatoxin exposure. *Mutat Res.* 2008;653(1-2):6-13. Epub 2008/05/10. doi: 10.1016/j.mrgentox.2008.01.012. PubMed PMID: 18467159.
71. Pineau P, Nagai H, Prigent S, Wei Y, Gyapay G, Weissenbach J, et al. Identification of three distinct regions of allelic deletions on the short arm of chromosome 8 in

hepatocellular carcinoma. *Oncogene*. 1999;18(20):3127-34. Epub 1999/05/26. doi: 10.1038/sj.onc.1202648. PubMed PMID: 10340384.

72. Marchio A, Pineau P, Meddeb M, Terris B, Tiollais P, Bernheim A, et al. Distinct chromosomal abnormality pattern in primary liver cancer of non-B, non-C patients. *Oncogene*. 2000;19(33):3733-8. Epub 2000/08/19. PubMed PMID: 10949927.

Figure Legends

Figure 1

A3DE dimerizes with itself, A3F and A3G. HEK293T cells were co-transfected by a GST-A3DE construct and an empty GST plasmid as negative control along with N-terminal 3xFLAG-tagged A3A, A3B, A3C, A3DE, A3F, A3G A3H as well as the A3BnA chimera. GST pull downs were followed by Western blotting with antibodies to 3xFLAG.

Figure 2

Confocal microscopy of A3DE with A3 proteins. Strong evidence for A3DE-V5, A3F-V5, A3G-V5 and the A3BnA-V5 chimera but scant evidence for A3B co-localizing with 3xFLAG-A3DE in HeLa cells 36 h after transfection. The scatter diagram and correlation coefficient (R) reveal significance of interaction.

Figure 3

A3DE_{C320Y} dimerizes and co-localizes with itself, A3DE, A3F and A3G. A) HEK293T cells were co-transfected by a 5' GST-A3DE, 5'GST-A3DE_{C320Y} constructs and an empty 5' GST as negative control along with N-terminal 3xFLAG-tagged A3DE, A3DE_{C320Y}, A3F and A3G. GST pull downs were followed by Western blotting with antibodies to 3xFLAG. B) Co-localization of A3DE_{C320Y}-V5, A3F-V5 and A3G-V5 with 3xFLAG-A3DE_{C320Y} in HeLa cells 36 h after transfection. The scatter diagram and correlation coefficient (R) reveal significance of interaction.

Figure 4

A3DE protects HBV DNA from A3 deamination in a dose-response manner. QT6 cells were transfected with pCayw and with A3F (A), A3G (B), A3B (C) and A3BnA (D) with increasing amounts of A3DE (0.25 µg; 0.50 µg; 0.75 µg) and complemented with empty plasmid up to 2 µg; pv, empty plasmid. The white line indicates the threshold between edited and unedited 3DPCR products. Asterisks refer to the samples cloned and sequenced. M: molecular weight markers.

Figure 5

A3DE interaction antagonizes A3F and A3G increasing HBV DNA levels. A) *APOBEC3* transcription profiling of transfected Huh7 cells in presence or absence of HBV. Data were normalized to the expression levels of *RPL13A* housekeeping reference gene. B) Huh7 cells were transfected with pCayw and pv or A3F or A3G with 0.25 µg and 0.75 µg of A3DE. At 7 days post transfection, total DNA from the supernatant was extracted and quantitative PCR was performed on HBV X gene. C) Huh7 cells were transfected with pCayw and siRNA control or A3DE siRNA knockdown. At 48h post transfection, total DNA from the supernatant was extracted and quantitative PCR was performed as described in B. * indicates a statistically significant difference between two observed values, ($p < 0.05$). D) Efficiency of A3DE siRNA. E) Huh7 cells were transfected by A3F or A3G and increasing concentrations of A3DE (DNA constant at 2 µg) and steady-state levels of A3F and A3G were determined by Western blotting at 24 h. The β -actin loading controls are shown below.

Figure 6

A3DEC320 is not functional and can be reverted by C320Y. (A) 3DPCR of the HBV X DNA. pCayw was transfected into the quail QT6, along with A3DE, A3DEC320Y or pv as negative control. At 72 hours post-transfection, total DNA was recovered and 3DPCR performed on a segment corresponding to the HBV X DNA as was described in (ref). The size of the 3DPCR fragment is 203 bp. Asterisks indicate the PCR products cloned and sequenced; M, molecular weight markers. B) Mutation matrices of hyperedited HBV sequences at two different temperatures. The numbers below the matrices indicate the number of bases sequenced. C) Frequency distribution of editing per clone at 86.5°C and 88.1°C. # indicates the number of clones sequenced. D) 5' Dinucleotide analysis of edited cytidine residues. Exp, expected values = dinucleotide composition. A χ^2 analysis indicates significantly different dinucleotide frequencies (*, $p < 0.05$).

Supporting Information

Supporting Table 1

Analysis of APOBEC3DEC320Y allele in patients and in cell lines.

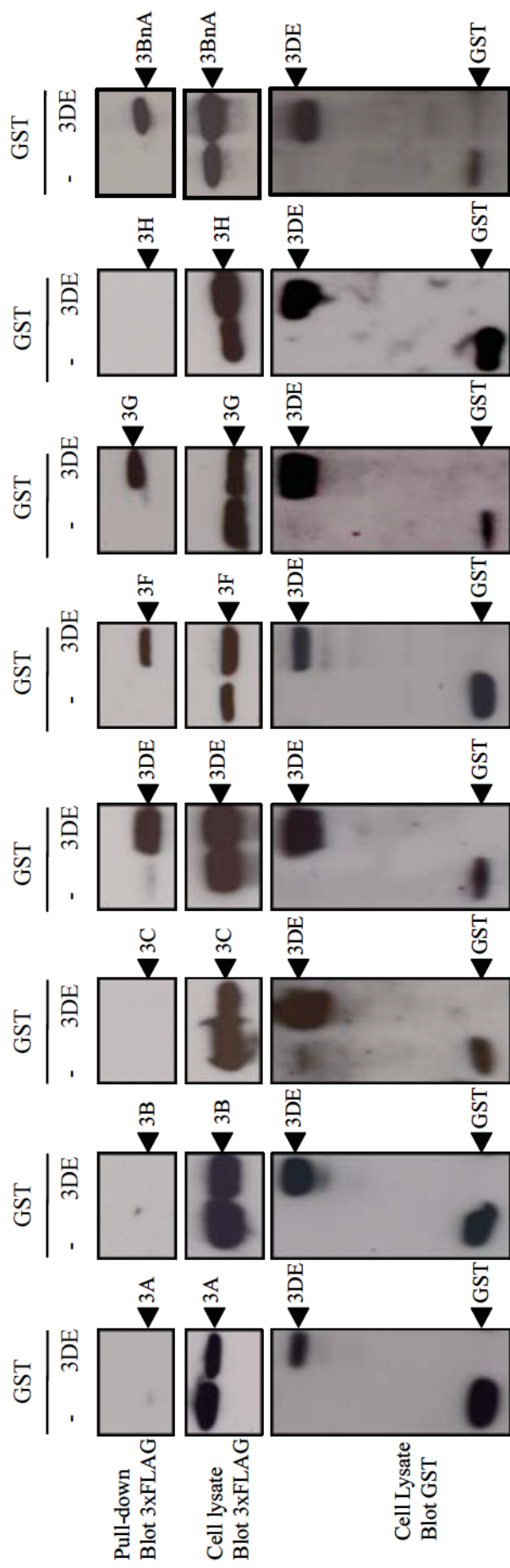


Figure 1

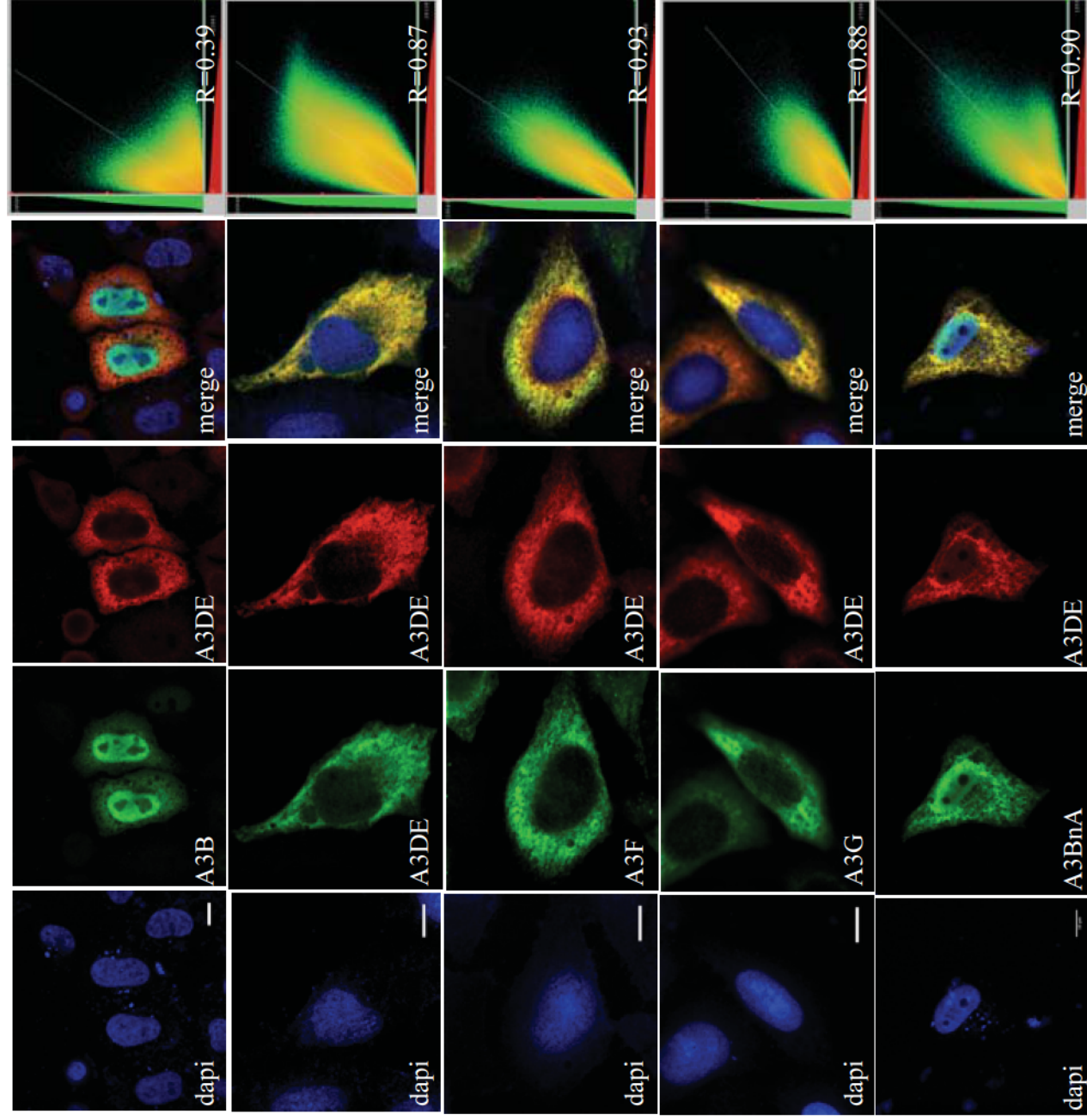


Figure 2

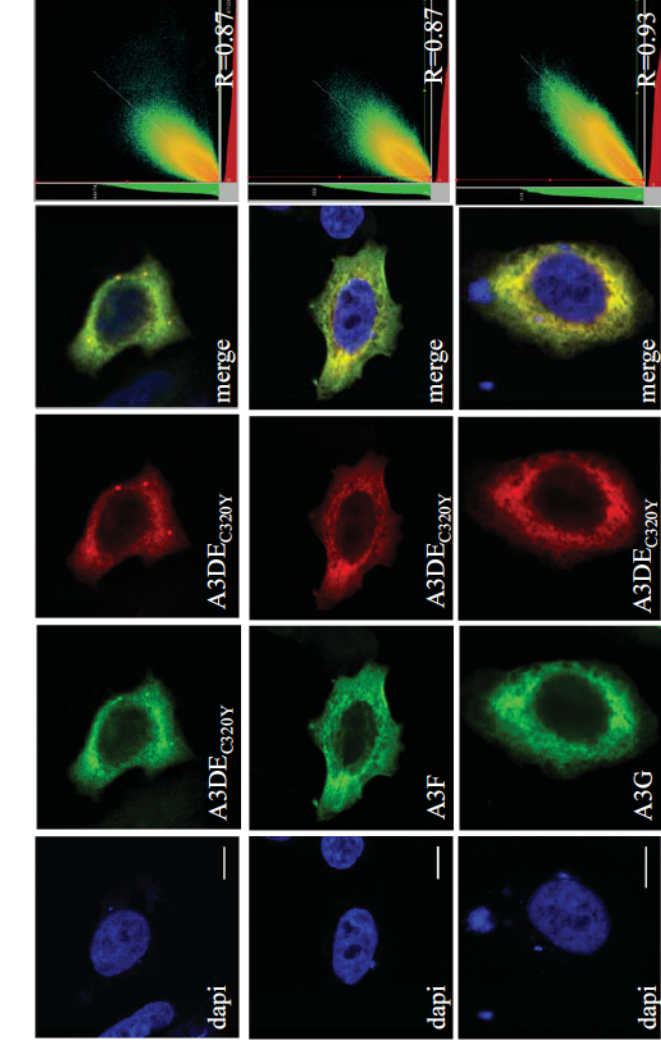
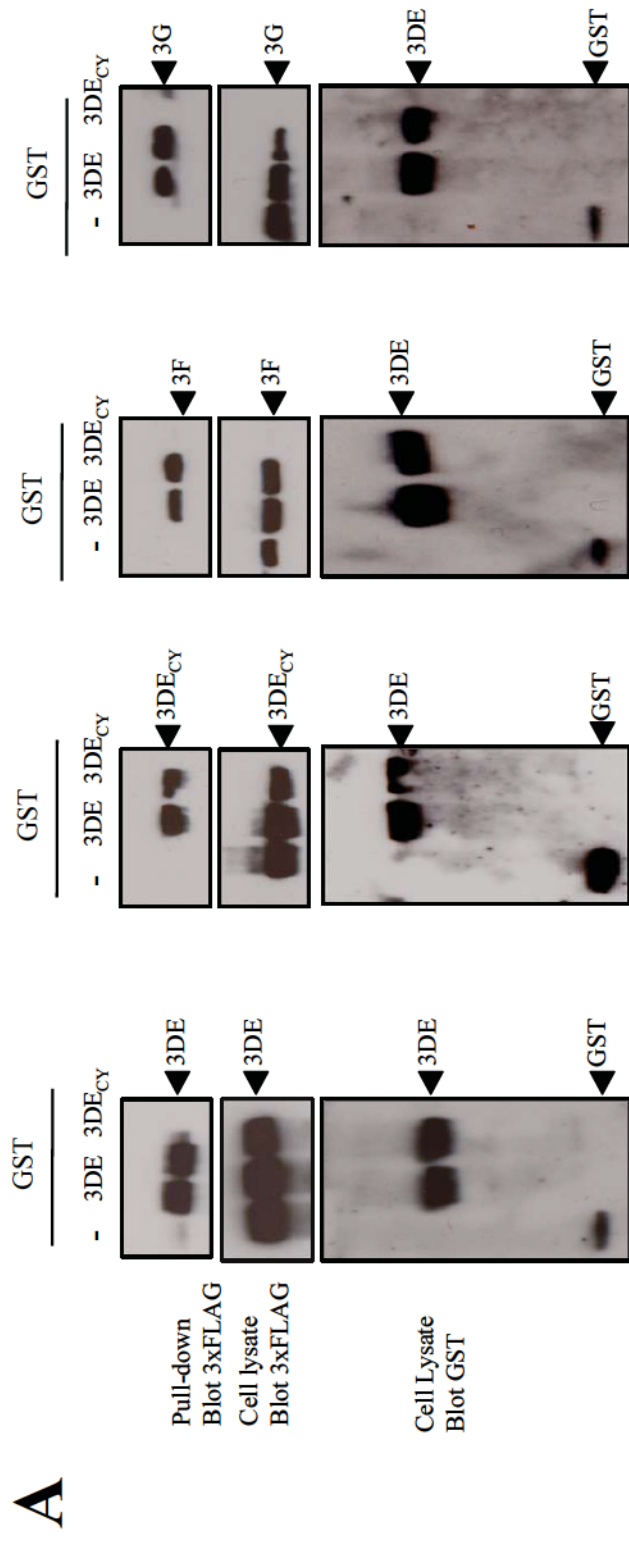


Figure 3

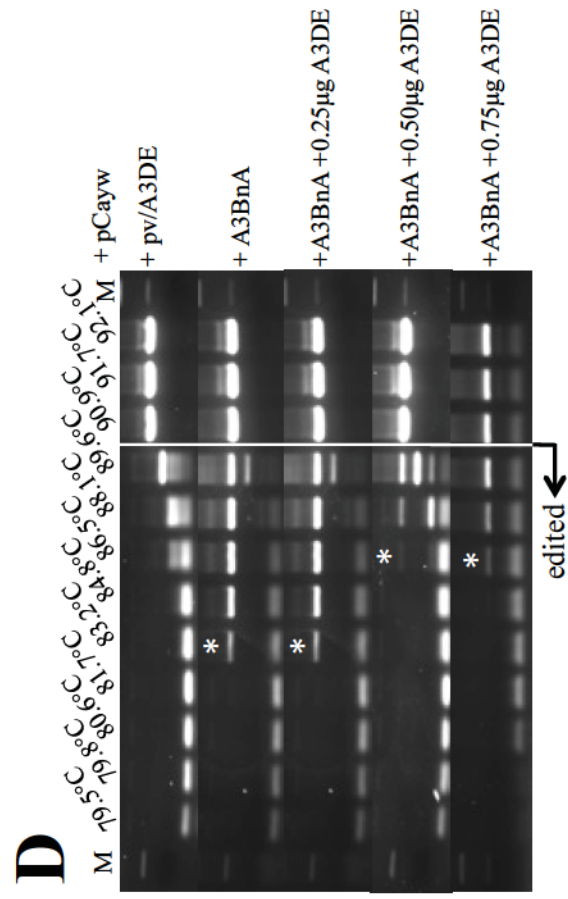
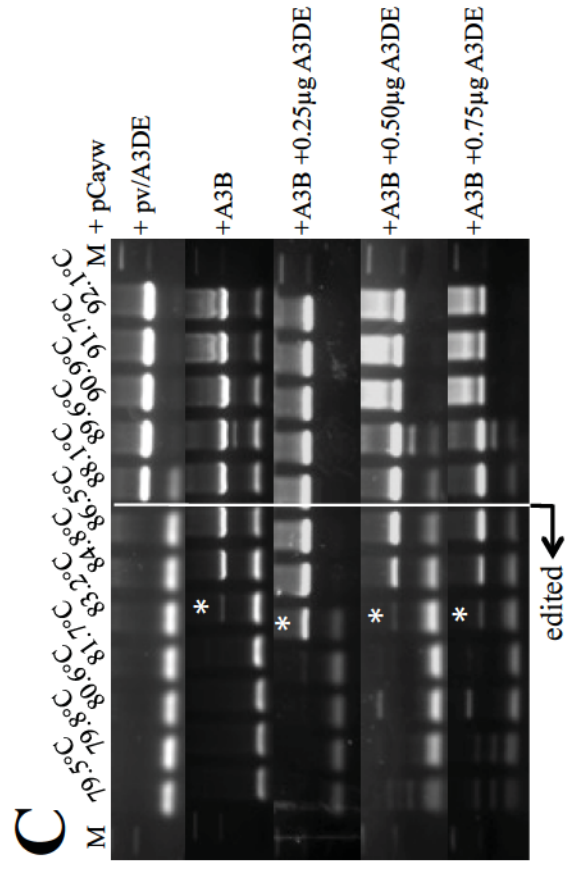
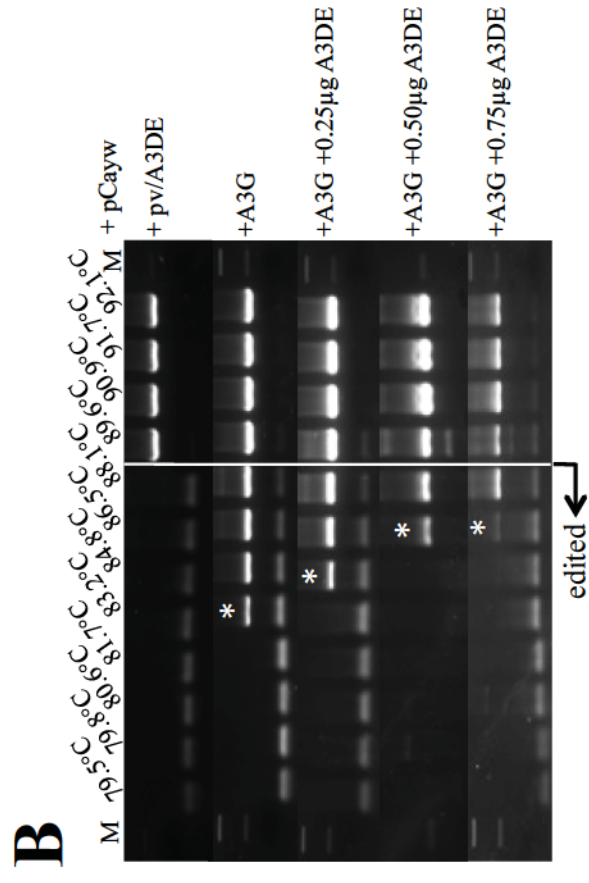
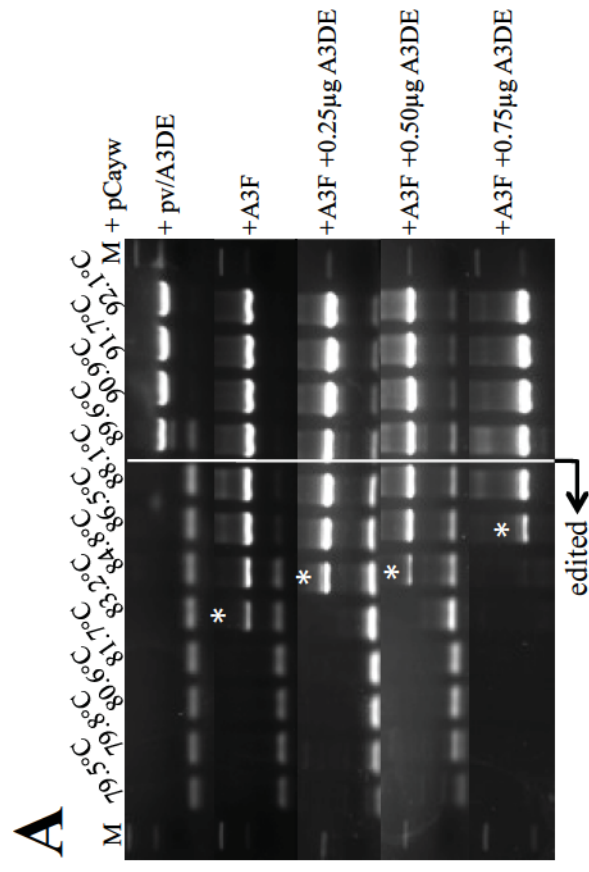


Figure 4

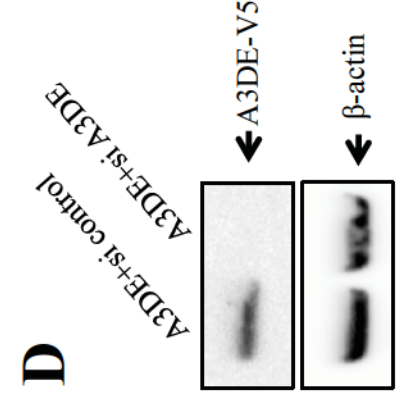
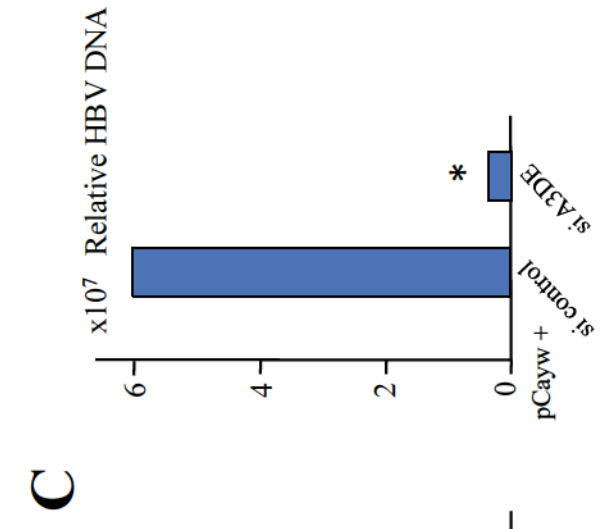
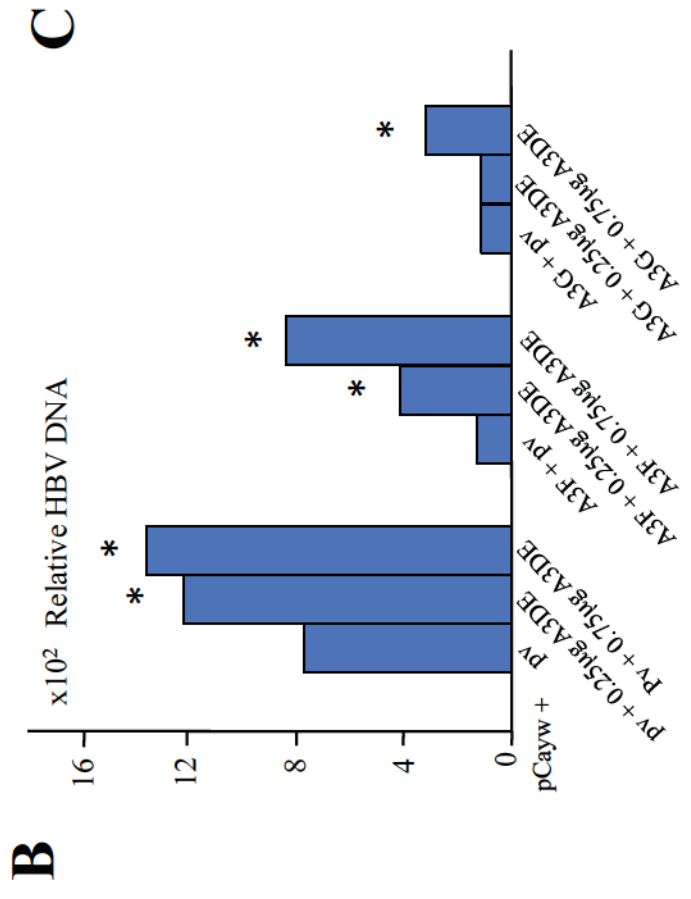
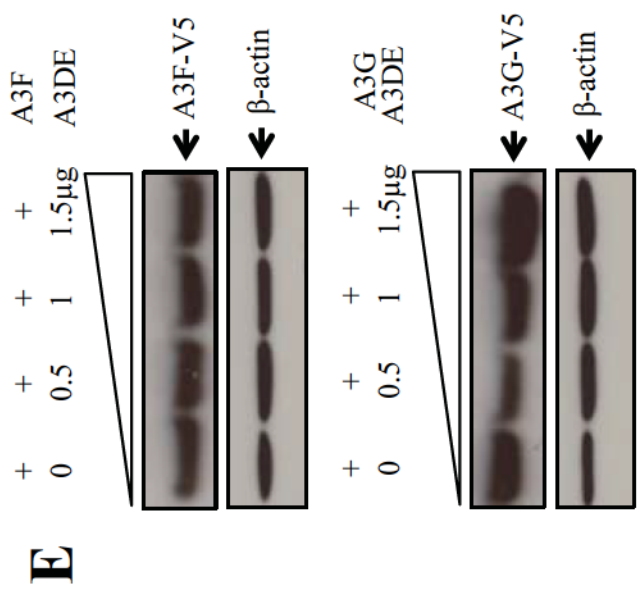
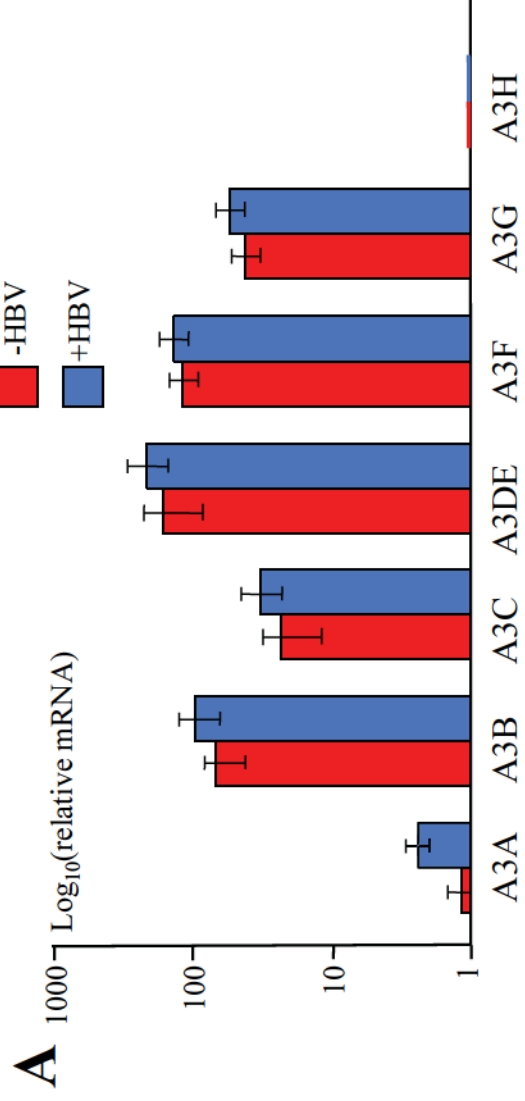
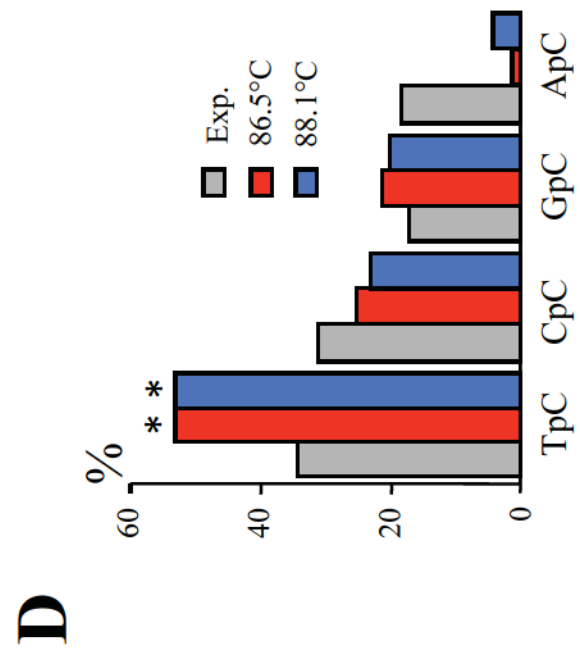
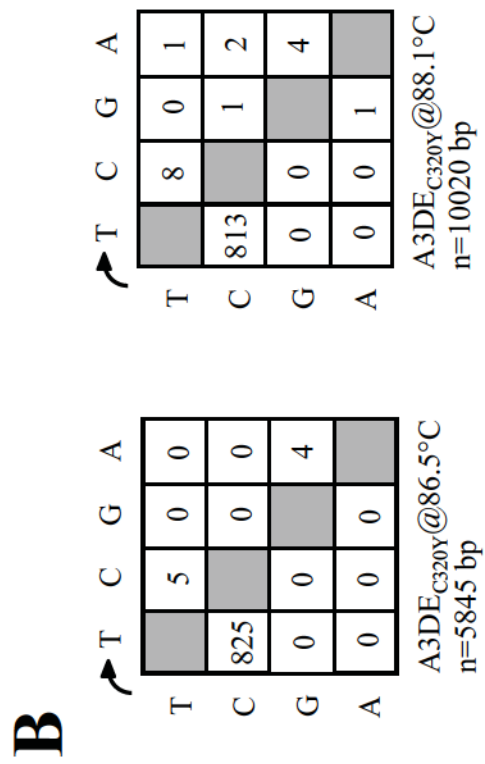
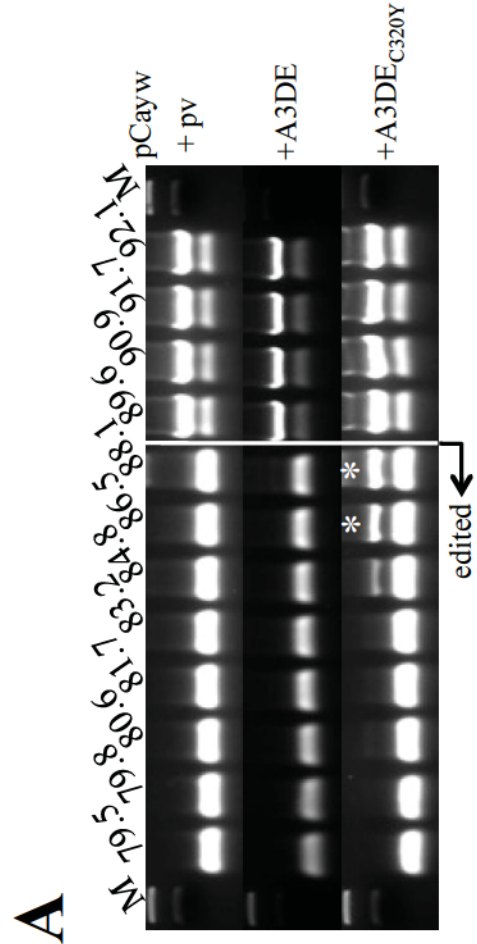


Figure 5



Number and % edit per seq.

Figure 6

Patients	
3 Healthy	Liver
10 HBV ⁺	Cirrhosis
10 HCV ⁺	Cirrhosis
11 HBV ⁺ HCV ⁺	Cirrhosis
10 Alcoholic	Cirrhosis
10 HBV ⁺ China	HCC
10 HBV ⁺ Italy	HCC
10 HBV ⁺ Morocco	HCC
10 HBV ⁺ Peru	HCC
10 HBV ⁺ Romania	HCC
10 HBV ⁺ Tunisia	HCC
Cell lines	
Huh7	HCC
OC4CH16	HCC
HLF	HCC
NUK-1	HCC
BEL9204	HCC
IHH	Immortalized hepatocyte
Fa2N	Immortalized hepatocyte
HHEmi	Primary hepatocytes
LSEC	Liver sinusoidal endothelial cell
HT29F	Colon adenocarcinoma
HCT116	Colorectal carcinoma
HaCaT	Immortalized keratinocyte
HEK293	Embryonic kidney

EMBO reports, 27,892 characters

SOUMIS

Molecular basis of the attenuated phenotype of human
APOBEC3B DNA mutator enzyme

Vincent Caval¹, Mohamed S. Bouzidi¹, Rodolphe Suspène¹, Hélène Laude¹,
Marie-Charlotte Dumargne², Anu Bashamboo², Thomas Krey³,
Jean-Pierre Vartanian¹ & Simon Wain-Hobson¹

¹Molecular Retrovirology Unit, Institut Pasteur, 28 rue du Dr. Roux
75724 Paris cedex 15, France,

²Human Developmental Genetics Unit, Institut Pasteur, 28 rue du Dr. Roux,
75724 Paris cedex 15, France,

&

³Structural Virology Unit, Institut Pasteur, 28 rue du Dr. Roux
75724 Paris cedex 15, France, ⁴CNRS UMR 3569, Paris, France

Correspondence to: SWH, Molecular Retrovirology Unit, Institut Pasteur, 28 rue du Dr. Roux,
75724 Paris cedex 15; T +33 1 45 68 88 21, F +33 1 45 68 88 74; simon.wain-hobson@pasteur.fr

Running title: Functional attenuation of APOBEC3B

Abstract

The human APOBEC3A and APOBEC3B genes (A3A and A3B) encode DNA mutator enzymes that deaminate cytidine and 5-methylcytidine residues in single-stranded DNA (ssDNA). They are important sources of mutations in many cancer genomes which show a preponderance of CG->TA transitions. Although both enzymes can hypermutate chromosomal DNA in an experimental setting, only A3A can induce double strand DNA breaks, even though they only differ by 9% at the protein level. Accordingly we sought the molecular basis underlying A3B attenuation through the generation of A3A-A3B chimeras and mutants. It transpires that the N-terminal domain facilitates A3B activity while a handful of substitutions in the catalytic C-terminal domain impacting ssDNA binding serve to attenuate A3B compared to A3A. Interestingly, functional attenuation is also observed for the rhesus monkey rhA3B enzyme compared to rhA3A indicating that this genotoxic dichotomy has been selected for and maintained for some 38 million years. Expression of all human ssDNA cytidine deaminase genes is absent in mature sperm indicting they contribute to somatic mutation and cancer but not human diversity.

Key words:

APOBEC3B / APOBEC3A / cancer / double stranded DNA breaks / genotoxicity

Introduction

Mutation-selection drives the oncogenic process [1, 2]. While replication errors and exogenous sources of DNA damage such as UV light, mutagens and free radicals have long been appreciated to be sources of somatic diversity driving oncogenesis [3], the contribution of two endogenous DNA mutator enzymes has only recently come to the fore [2, 4-9]. The APOBEC3A (A3A) and APOBEC3B (A3B) enzymes are cytidine deaminases, converting cytidine residues into uridine bases in single stranded DNA (ssDNA) [6, 9-11]. The non-canonical uracil bases are removed by uracil N-glycosylase (UNG) generating abasic sites in ssDNA, which in turn are cleaved by apurinic/apyrimidic endonucleases, APE1 and APE2. Together, the three enzymes effectively function as a cytidine endonuclease, promoting apoptosis or DNA repair depending on the extent of DNA damage.

Although the A3 enzymes were initially described as innate immune restriction factors for many DNA viruses, retroviruses and retroelements, A3A and A3B have emerged as sources of the large number of CG->TA somatic mutations in cancer genomes and nuclear DNA (nuDNA) in experimental settings [2, 4-9]. A3 editing leaves a tell tale 5'TpC mutational signature, where C is the edited base, which shows up in many cancer genomes [2, 4, 8, 12, 13]. Epidemiological data links A3A to the development of breast, ovarian and hepatitis B virus associated liver cancer [7, 14-16]. While both A3A and A3B can deaminate 5-methylcytidine residues [6, 17], the two enzymes are not orthologous for A3B overexpression fails to produce double stranded DNA breaks (DSBs), unlike A3A [6].

A3A is a 26 kDa enzyme with a single zinc finger catalytic domain [18]. By contrast APOBEC3B is a 46 kDa enzyme made up of two zinc finger domains, the inactive N-terminal domain (A3Bn) domain being more closely related to the N-terminal domains of APOBEC3DE, F and G, while the catalytically active C-terminal domain, A3Bc, arose from duplication of the *A3A* gene [19-21]. Because of this, A3A and A3Bc are 91% identical at the protein level [6]. To explore the molecular basis of A3B attenuated DSB phenotype compared to A3A, a large number of chimeras and mutants were generated and their genotoxicity assessed. The N-terminal domain of A3B and at least three residues involved in DNA binding by the A3Bc catalytic domain contribute to A3B attenuation. The rhesus monkey A3B enzyme is equally attenuated compared to rhA3A indicating that the trade off between A3A and A3B has been maintained for at least 38 million years.

Results

Both N- and C-terminal A3B domains impact editing

In order to confer DSB formation, extensive mutagenesis of the catalytic C-terminal domain of A3B was first made. However, none of a series of 12 mutants encoding 1-3 residue substitutions as well as an A3B specific three-residue deletion (Fig EV1A) was able to rescue the DSB formation phenotype as assessed by FACS analysis of γ H2AX staining, suggesting that A3B attenuation involved multiple residues. To attack the problem from a different angle, a chimeric construct was used whereby the A3Bc domain was replaced by A3A (A3BnA). A second construct comprising only A3Bc was made using the equivalent of the second Met¹ in the A3A sequence, which leads to the functional A3Ap2 isoform (Fig 1A) [22, 23]. A3BnA was well expressed while steady state A3Bc concentrations were lower than for A3A (Fig 1B). Interestingly, if A3Bc displayed a nucleo-cytoplasmic distribution similar to A3A, the A3BnA chimera was mostly found in the nucleus although retained some cytoplasmic distribution, suggesting that residues in both A3B domains are required for A3B localization which is strictly nuclear (Fig 1C).

To evaluate deaminase activity, A3BnA and A3Bc-V5 tagged plasmids were transfected into 293T cells and cell lysates used in an *in vitro* FRET based assay where C to U deamination of a TAM-FAM labeled DNA oligonucleotide allows fluorescence detection after uracil-DNA glycolase (UNG) cleavage [6, 11]. While A3BnA displayed cytidine deaminase activity equal to A3A and A3B, only very weak activity was observed with A3Bc (Fig 1D). As HBV genome editing has proven to be a very sensitive assay for A3 activity *in vivo* [24], QT6 cells which do not show an endogenous editing background were transfected with A3B plasmids along with the pCayw HBV infectious molecular clone and HBV editing analyzed by 3DPCR. This technique allows recovery of AU rich A3-edited DNA at lower PCR denaturation temperatures compared to unedited DNA [25]. For all constructs, HBV DNA was recovered at temperatures below the limiting denaturation temperature of T_d=89.0°C corresponding to unedited DNA obtained with inactive A3BE255Q or A3AC106S mutants (Fig 1E). Editing was confirmed by cloning and sequencing of 3DPCR products. If A3BnA activity was on a par with A3A and A3B, A3Bc again proved to be less active (Fig 1E).

To assess the editing of nuclear DNA (nuDNA), 293T-UGI cells were used [6, 9]. Using cellular *TP53* specific 3DPCR, edited nuDNA was recovered following transfection. As

previously described, A3A consistently outperformed A3B with Tds down to 84.6°C compared to 85.7-86.6°C [6]. A3BnA was able to hypermutate nuDNA, just like A3A and A3B while A3Bc did not (Fig 1F). As 3DPCR is not a quantitative technique, genotoxicity was quantitated for DSB formation, it being a more macroscopic indicator of A3 damage. Both A3B and A3Bc did not make detectable DSBs. By contrast A3BnA transfection resulted in significant levels of γ H2AX phosphorylation albeit ~50% compared to the A3Ap1 and A3Ap2 controls indicating that A3Bn can attenuate A3A activity (Fig 2A, 2C). Given these results, A3BnA mutants can be used for loss of function studies while A3Bc can be used to screen for gain of function mutants.

As all 18 residues (9%) distinguishing A3Ap2 from A3Bc were located in three regions referred to as mutation clusters 1, 2 and 3 (Fig 1A), six A3BnA and six A3Bc chimeras were made (Fig 2A). Expression levels were comparable although there were subtle differences in nuclear localization, which was completely nuclear for A3BnA1, A3BnA3 and A3BnA1-3, suggesting that some A3B residues present in cluster 2 are necessary for complete nuclear addressing (Fig 2B and Fig EV1B). For the A3BnA derivatives exchange of almost any region of the C-terminal A3A domain resulted in loss of DSB formation (Fig 2C). By contrast, A3Bc mutation cluster 1 (A3Bc1) restored DSB formation to levels comparable to A3A and A3Ap2. The second cluster of mutations (A3Bc2) restored DSB formation but not quite as well, while A3Bc3 had a weak but significant effect (Fig 2C). Cluster 2 and 3 mutations combined were comparable to A3Bc1 and positive controls, A3Ap1 or A3Ap2.

Molecular determinants of A3Bc attenuation

To pinpoint gain of function mutations in A3Bc1 and A3Bc2 constructs conferring DSB formation, smaller combinations of A3Ap2 residues were introduced into the A3Bc coding plasmid (Fig 3A). Of all the constructs tested, only the RQ212HK, and to a lesser degree the DT196HI mutant, were found to induce significant DSB breaks (Fig 3A). The combined quadruple DT196HI+RQ212HK mutant recovered further DSB activity, more so than the sum of the two, but still only ~50% DSB formation compared to A3Ap1 or A3Ap2 (Fig 3B), suggesting that other substitutions in mutant cluster 1 that do not score positive alone, when combined with DT196HI and RQ212HK contribute to genotoxicity. Single residue mutagenesis showed that the Q213K substitution was the most important with D196H and T197I scoring just above background levels (Fig 3B). Interestingly, reciprocal mutagenesis of

A3Ap2 showed that the HK29RQ (A3A numbering, Fig 1A) construct attenuated DSB formation more than the HI16DT exchange mirroring the changes on the A3Bc background (Fig 3C). An NMR analysis of ssDNA binding to A3A identified K30 and to a lesser extent I17 residues (respectively Q213 and T197 on A3B, Fig 1A) as undergoing chemical shift changes, stressing their participation in substrate binding. These residues participate in binding the ssDNA backbone (Fig 3D, 3E).

Conserved A3Bc attenuation

A3A enzymes are orthologous across placental mammals [26]. By contrast A3B is a feature of primate genomes at least from *Cercopithecus* monkeys to man. Phylogenic analysis of A3A/A3Bc pairs from primates, revealed that rhesus macaque rhA3A and rhA3Bc clustered together and apart from the A3A or A3Bc groups defined by the human, chimp and bonobo enzymes suggesting some recent gene conversion (Fig EV2A). Of the 22 residues (11%) distinguishing rhA3A and rhA3Bc only six sites were also variable between human A3A and A3Bc (Fig 4A). To see if the same genotoxic dichotomy existed for the rhesus enzymes, four rhA3A and rhA3B constructs were made, found to be well expressed and were localized comparably to their human counterparts (Fig 4B, 4C). RhA3B and rhA3Bc were unable to produce DSBs unlike rhA3A and rhA3Ap2, so paralleling their human counterparts (Fig 4D). Of the macaque A3A/B residues corresponding to human HT16DT and K213Q only residue was variable and then as a known D/N polymorphism (Fig 4A, boxed) [26]. Overall the substitutions in both human and macaque A3Bc domains make them more acidic compared to A3A that could reduce the overall affinity for negatively charged ssDNA. Interestingly, this observation extended to every available A3Bc sequence, consistently being the most acidic of all A3 domains (Fig EV2B).

In view of the attenuated yet conserved A3B phenotype compared to A3A, we wondered if it could contribute to human variation by looking at the RNAseq transcriptome in three purified mature sperm samples (Fig EV3). *APOBEC* transcript frequencies were of the order of $4-16 \times 10^{-5}$ of total reads compared to the sperm specific protamine *PRM1* gene which was correctly expressed. No *A3B* transcripts were identified. *APOBEC4* encodes a protein with clear identity to other APOBEC enzymes yet is devoid of catalytic activity. *APOBEC4* levels were a little higher ($3-9 \times 10^{-4}$ of total reads) but still very low compared to *PRM1* that contrasts with abundant *APOBEC4* expression in human testis [27]. As ~60% of sperm show motility

defects, the very small number of APOBEC transcripts, which were not consistent between samples, most likely means that they are without consequence in sperm.

Discussion

While it is now appreciated that A3A and A3B are both endogenous human DNA mutators, their phenotypes are different and non-overlapping. A3A is by far the more genotoxic of the two (Fig 1F and 2C), which is the more surprising given that only of a small fraction of A3A localizes to the nucleus. A3A expression is tightly regulated, with steady state levels of A3A kept very low in all most cells. However, A3A expression is very sensitive to IFN α , IFN γ and PMA [11, 22, 28, 29]. By contrast A3B is strictly nuclear, expressed in most tissues and is insensitive to IFN α . However, if both enzymes are pro-apoptotic and can introduce mutations in nuDNA as detected by highly sensitive 3DPCR [6, 9, 22], A3B consistently failed to generate DSBs above background levels [6].

The underlying mechanisms for this dichotomy are several. First, the N-terminal A3Bn domain plays a major role in attenuating A3B function for the introduction of cluster 1, 2 and 3 residues from A3A are fail to rescue a DSB formation phenotype as they do in a A3Bc backbone (Fig 2C). In addition when fused to the N-terminus of A3A it attenuated A3A function (Fig 2C). Second, exchange of cluster 1 and 2 residues from A3A clearly conferred DSB capacity on the chimeric A3Bc constructs (Fig 2C). Of the eight substitutions and the deletion present in cluster 1, D196H, T197H and Q213K appeared crucial for DSB formation, the latter being the most important (Figure 3B). Nonetheless, the quadruple A3Bc mutant DT196HI+RQ212HK only recovered ~50% of DSB activity compared to A3A (Fig 3A) suggesting that other cluster 1 substitutions are involved whose individual contributions do not score above background. The effects of residues defining cluster 2 are equally subtle and context dependent, since replacement of three residues alone in A3Bc had no effect on DSB induction (Fig 3A). Cluster 2 and 3 mutations together (A3Bc2-3) performed as well as cluster 1 mutations alone (A3Bc1) and A3Ap1 and A3Ap2 indicating a certain degree of overlap and degeneracy.

Interestingly, the genotoxic dichotomy between A3A and A3B was also apparent for the rhesus macaque enzymes. Although the 22 residues distinguishing the catalytic domains of

two enzymes were more widely distributed across the sequence (Fig 4A). Of the three residues distinguishing human A3A and A3B only D16N was variable. However, as this is a polymorphic between equally functional rhA3A enzymes [26] it cannot explain the difference. As the rhesus A3A and A3B sequences cluster together and not with the human, chimp and bonobo clusters of A3A or A3B sequences perhaps a more recent gene conversion event other residues underwent selection to maintain the A3A/A3B genotoxic dichotomy.

A3B deletion haplotype frequencies of 30-40% in SE Asia [30] not only reinforces the observation that the A3B phenotype is weak but also that is not required for survival and fertility. The RNAseq analysis of purified mature human sperm shows that endogenous human cytidine deaminases are principally active in the soma and may not be a source of human diversity. While recently evolved human genomes show a penchant for 5'TCC -> 5'TTC transitions their frequency was the same for Chinese and Yoruba populations which differ considerably in terms of the A3B deletion allele [30] again arguing against A3A and A3B playing a role in evolution of the human genome [31]. Nonetheless, the conservation of attenuated A3B genotoxicity compared to A3A across primates argues for an essential role in somatic cells, albeit a subtle one.

Materials and Methods

Plasmids

APOBEC3A, APOBEC3Ap2 and APOBEC3B expression plasmids and catalytic mutants have been previously described [6, 9, 22] and were used to generate chimeras by PCR (Table EV1) or site directed mutagenesis (GeneArt® Site-Directed Mutagenesis System, Life Technologies) (Table EV2). To overcome toxicity in *E. coli*, A3BnA and rhesus APOBEC3B (RhA3B) plasmids were synthesized retaining intron 7, and subsequently cloned into pcDNA3.1D/V5-His-TOPO vector (Life Technologies). Rhesus-A3A expression plasmid was previously described and used to generate Rhesus-A3Ap2 plasmid using PCR. All constructs were grown in *E. coli* TOP10 cells (Life Technologies) and verified by sequencing.

Cells

Quail QT6 embryo fibroblast cells were maintained in HAM's F40 medium (Eurobio), supplemented with 1% chicken serum, 10% FCS, 5% tryptose phosphate, 2 mM L-glutamine,

50 U/ml penicillin and 50 mg/ml streptomycin. Human HeLa cells and 293T-UGI cells stably expressing *Bacillus subtilis* phage uracil-DNA glycosylase inhibitor (UGI) were maintained in DMEM glutamax medium (Life Technologies) supplemented with 10% FCS, 50 U/ml penicillin and 50 mg/ml streptomycin.

Transfections

One million QT6 cells were co-transfected with 0.5 µg of pCayw HBV coding plasmid and 1.5 µg APOBEC3 expression plasmids using JetPrime (Polyplus) following manufacturer's recommendations and harvested 48 hours post-transfection. For single plasmid transfections, 8×10^5 of HeLa, 293T-UGI, cells were transfected using 2 µg APOBEC3 expression plasmids using JetPrime (Polyplus) following manufacturer's recommendations and harvested 48 hours post-transfection. For immunofluorescence labeling, 5×10^4 HeLa cells grown on chamber slides (LabTek) were transfected with either 1 µg APOBEC3 expression plasmids using Fugene HD (Roche) following manufacturer's recommendations.

Western blotting

Transfected cells were resuspended in lysis buffer (0.5% Nonidet P-40, 20 mM Tris-HCl pH7.4, 120 mM NaCl, and 1 mM EDTA) supplemented with Complete Protease Inhibitor Mixture (Roche Applied Science). Cell lysates were incubated on ice for 20 min and then clarified by centrifugation at 14,000 x g for 30 min. Western blot analysis on cell lysates was carried out according to standard procedures. After blocking, membranes were probed with 1/5000 diluted mouse monoclonal antibody specific for the V5 epitope (Life Technologies) in PSB-0.1% Tween 5% dry milk applied overnight. After PBS-Tween washings and incubation with an anti-mouse IgG horseradish peroxidase-coupled secondary antibody (Amersham), the membrane was revealed by enhanced chemiluminescence (Pierce). β -Actin was used as a loading control using 1/2000 diluted mouse monoclonal antibody specific for β -actin (Sigma).

Immunofluorescence

After PBS washings, transfected HeLa cells grown on chamber slides were fixed with 4% PFA for 15 min. After PBS washing cells were incubated in 50/50 acetone/methanol for 20 minutes. Mouse monoclonal anti-V5 antibody (Life Technologies), and rabbit monoclonal anti-FLAG (Sigma Aldrich) were incubated at 1/200 for 1 h at room temperature, followed by incubation with a mouse specific Alexa-488 goat antibody and rabbit specific alexa-555 conjugated donkey antibody 1 h at room temperature in the dark. After washing, slides were

mounted with Vectashield imaging medium containing DAPI (Vector Laboratories). Imaging was performed using Leica SP5 confocal microscope.

***In vitro* deamination assay**

At 72 hours after transfection, APOBEC3 transfected 293T cells were extensively wash with PBS and mechanically harvested. Total proteins were extracted using specific lysis buffer (25 mM HEPES pH7.4, 10% glycerol, 150 mM NaCl, 0.5% Triton X-100, 1 mM EDTA, 1 mM MgCl₂, 1 mM ZnCl₂) supplemented with protease inhibitors, and submitted to sonication. Deaminase activity was assessed by incubating whole cell lysates with 1pmole DNA oligonucleotide 5'-(6-FAM)-AAATTCTAATAGATAATGTGA-(TAMRA)-3' in presence of 0.4 unit UDG (NEB) in a 20 mM Tris-HCl, 1 mM DTT, 1 mM EDTA reaction buffer. After 2 hours incubation at 37°C, generated abasic sites were cleaved by heating 2 minutes at 95°C, and endpoint fluorescence were measured using realplex² Mastercycler (BioRad) with FAM setting and background fluorescence obtained with mock-transfected cells set as negative control. Results are normalized to the quantity of protein using Pierce BCA Protein Assay Kit (Thermo Scientific).

DNA extraction and 3DPCR amplification

Total DNA from transfected cells was extracted using the MasterPureTM complete DNA and RNA purification kit (Epicentre) and resuspended in 30 µL sterile water. All amplifications were performed using first-round standard PCR followed by nested 3DPCR [9, 24]. PCR was performed with 1 U Taq DNA polymerase (Eurobio) per reaction.

FACS analysis of double strand breaks

At 48 hours after transfection, cells were washed with PBS, fixed in 2-4% ice-cold paraformaldehyde (Electron Microscopy Sciences) for 10 minutes and permeabilized in 90% ice-cold methanol (Sigma) for 30 minutes. After washing with PBS, cells were incubated with 1:200 diluted mouse anti-V5 antibody (Life Technologies) in PBS-BSA 0.5% for 1 hour. After PBS washings incubation with 1:500 diluted Alexa Fluor 633 F(ab')₂ fragment of goat anti-mouse IgG (H+L) (Life Technologies) was performed for 45 minutes. DNA double strand breaks were analyzed by staining for 1 hour with 1:50 diluted Alexa Fluor 488-conjugated rabbit monoclonal anti-γH2AX (20E3) antibody (Cell Signaling). All incubation steps were performed on ice. Stained samples were analyzed on a FACSCalibur using

CellQuest Pro (BD Biosciences, version 5.2) and data were analyzed with FlowJo software (Tree Star Inc. version 8.7.1).

Sperm RNaseq

Spermatozoa from three individuals were selected and separated from somatic cells by centrifuging 1 to 2 ml of semen through a two-layer (45% and 90%) gradient of Percoll at 400x g for 20 min. The fraction in the 90% layer was washed in 2 ml of Tyrode (Eurobio) and centrifuged at 400x g for 10 min. The pellet was collected with care to avoid somatic cell contamination. The pellet was washed twice in 1x PBS and RNA was isolated using RNeasy® mini kit (Qiagen). Genomic DNA contamination was eliminated first by performing on-column DNase Digestion with the RNase-Free DNase Set (Qiagen). Subsequently, the purified RNA samples were subjected to RNase-free DNase treatment (TURBO™ DNase, Ambion Inc).

RNaseq analysis was performed using SOLiD™ Whole Transcriptome analysis kit for whole transcriptome libraries (Applied biosystems™) following the manufacturer's instruction. Briefly, total RNA isolated from spermatozoa was fragmented, purified, linked to adaptors and reverse transcribed to cDNA. This double stranded template was size selected on TBE-urea gel and DNA fragments between 100-200 bp in length were used to construct a SOLiD™ whole transcriptome library (Applied Biosystems SOLiD™ 3 System Templated Bead Preparation Guide) that was subsequently sequenced.

gColor Space sequence reads were mapped and quantified using the ABI Bioscope 1.3 Whole Transcriptome Analysis pipeline, the Integromics SeqSolve analysis software and a set of *ad hoc* perl scripts. The reference sequence datasets considered were the UCSC hg19 repeat-masked genome sequence and the NCBI RefSeq transcriptome dataset (September 2010). A purity score (calculated with a perl script) was assessed for each sample in order to evaluate the relative proportion of transcripts originating from cell types other than spermatozoa as described elsewhere [32]. Purity was defined as the ratio of each of the potential "contaminants" as a function of the average intensity for the sperm-specific *PRM1* and *PRM2* transcripts.

Structure analysis

Figures were prepared with Pymol (<http://www.pymol.org>). The calculation of pKa values of individual residues was performed using PROPKA 3.0 [33] on the first three models in their order of appearance in the coordinate file (PDB 2M65). Further representations used the first model of the coordinate file. The electrostatic potential, represented in the figures ramp-colored from red (negative) to blue (positive) through white (neutral), was calculated using APBS [34] and contoured on a scale ranging from -5 to 5 kT/e. Mutagenesis was performed using Pymol choosing the most common rotamer for this residue that did not cause clashes.

Acknowledgements

We would like to thank Milana Shapira for plasmid constructions and the Institut Pasteur's Plateforme d'Imagerie Dynamique for help with confocal microscopy. This work was supported by grants from Institut Pasteur, INCa and La Ligue Nationale contre le Cancer. VC was supported postdoctoral grant from OSEO.

Author Contributions

VC, JPV & SWH designed the study; VC, MSB, RS, HL performed the A3 experiments, MCD the RNAseq analysis and TK APOBEC3A modeling. VC & SWH wrote the paper.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Bignell GR *et al* (2010) Signatures of mutation and selection in the cancer genome. *Nature* **463**: 893-898
2. Greenman C *et al* (2007) Patterns of somatic mutation in human cancer genomes. *Nature* **446**: 153-158
3. Lengauer C, Kinzler KW, Vogelstein B (1998) Genetic instabilities in human cancers. *Nature* **396**: 643-649
4. Alexandrov LB *et al* (2013) Signatures of mutational processes in human cancer. *Nature* **500**: 415-421
5. Burns MB *et al* (2013) APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* **494**: 366-370
6. Caval V, Suspene R, Shapira M, Vartanian JP, Wain-Hobson S (2014) A prevalent cancer susceptibility APOBEC3A hybrid allele bearing APOBEC3B 3'UTR enhances chromosomal DNA damage. *Nat Commun* **5**: 5129
7. Komatsu A, Nagasaki K, Fujimori M, Amano J, Miki Y (2008) Identification of novel deletion polymorphisms in breast cancer. *Int J Oncol* **33**: 261-270
8. Nik-Zainal S *et al* (2012) Mutational Processes Molding the Genomes of 21 Breast Cancers. *Cell* **149**: 979-993
9. Suspène R *et al* (2011) Somatic hypermutation of human mitochondrial and nuclear DNA by APOBEC3 cytidine deaminases, a pathway for DNA catabolism. *Proc Natl Acad Sci USA* **108**: 4858-4863
10. Kostrzak A, Henry M, Demoyen PL, Wain-Hobson S, Vartanian JP (2014) APOBEC3A catabolism of electroporated plasmid DNA in mouse muscle. *Gene Ther* **22**:96-103
11. Stenglein MD, Burns MB, Li M, Lengyel J, Harris RS (2010) APOBEC3 proteins mediate the clearance of foreign DNA from human cells. *Nat Struct Mol Biol* **17**: 222-229
12. Roberts SA *et al* (2013) An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat Genet* **45**: 970-976
13. Stephens PJ *et al* (2009) Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* **462**: 1005-1010
14. Long J *et al* (2013) A Common Deletion in the APOBEC3 Genes and Breast Cancer Risk. *J Natl Cancer Inst* **105**: 573-579
15. Xuan D, Li G, Cai Q, Deming-Halverson S, Shrubsole MJ, Shu XO, Kelley MC, Zheng W, Long J (2013) APOBEC3 deletion polymorphism is associated with breast cancer risk among women of European ancestry. *Carcinogenesis* **34**: 2240-2243

16. Zhang T *et al* (2012) Evidence of associations of APOBEC3B gene deletion with susceptibility to persistent HBV infection and hepatocellular carcinoma. *Hum Mol Genet* **22**: 1262-1269
17. Suspène R, Aynaud MM, Vartanian JP, Wain-Hobson S (2013) Efficient deamination of 5-methylcytidine and 5-substituted cytidine residues in DNA by human APOBEC3A cytidine deaminase. *PLoS ONE* **8**: e63461
18. Jarmuz A, Chester A, Bayliss J, Gisbourne J, Dunham I, Scott J, Navaratnam N (2002) An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics* **79**: 285-296
19. Henry M, Guetard D, Suspene R, Rusniok C, Wain-Hobson S, Vartanian JP (2009) Genetic editing of HBV DNA by monodomain human APOBEC3 cytidine deaminases and the recombinant nature of APOBEC3G. *PLoS ONE* **4**: e4277
20. LaRue RS *et al* (2008) The artiodactyl APOBEC3 innate immune repertoire shows evidence for a multi-functional domain organization that existed in the ancestor of placental mammals. *BMC Mol Biol* **9**: 104
21. Munk C, Willemsen A, Bravo IG (2012) An ancient history of gene duplications, fusions and losses in the evolution of APOBEC3 mutators in mammals. *BMC Evol Biol* **12**: 71
22. Mussil B, Suspène R, Aynaud MM, Gauvrit A, Vartanian JP, Wain-Hobson S (2013) Human APOBEC3A isoforms translocate to the nucleus and induce DNA double strand breaks leading to cell stress and death. *PLoS ONE* **8**: e73641
23. Thielen BK, McNevin JP, McElrath MJ, Hunt BV, Klein KC, Lingappa JR (2010) Innate immune signaling induces high levels of TC-specific deaminase activity in primary monocyte-derived cells through expression of APOBEC3A isoforms. *J Biol Chem* **285**: 27753-27766
24. Vartanian JP *et al* (2010) Massive APOBEC3 editing of hepatitis B viral DNA in cirrhosis. *PLoS Pathog* **6**: e1000928
25. Suspène R, Henry M, Guillot S, Wain-Hobson S, Vartanian JP (2005) Recovery of APOBEC3-edited human immunodeficiency virus G->A hypermutants by differential DNA denaturation PCR. *J Gen Virol* **86**: 125-129
26. Caval V, Suspene R, Vartanian JP, Wain-Hobson S (2014) Orthologous mammalian APOBEC3A cytidine deaminases hypermutate nuclear DNA. *Mol Biol Evol* **31**: 330-340
27. Rogozin IB, Basu MK, Jordan IK, Pavlov YI, Koonin EV (2005) APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases predicted by computational analysis. *Cell Cycle* **4**: 1281-1285
28. Aynaud MM, Suspène R, Vidalain PO, Mussil B, Guétard D, Tangy F, Wain-Hobson S, Vartanian JP (2012) Human Tribbles 3 protects nuclear DNA from cytidine deamination by APOBEC3A. *J Biol Chem* **287**: 39182-39192

29. Koning FA, Newman EN, Kim EY, Kunstman KJ, Wolinsky SM, Malim MH (2009) Defining APOBEC3 expression patterns in human tissues and hematopoietic cell subsets. *J Virol* **83**: 9474-9485
30. Kidd JM, Newman TL, Tuzun E, Kaul R, Eichler EE (2007) Population stratification of a common APOBEC gene deletion polymorphism. *PLoS Genet* **3**: e63
31. Harris K (2015) Evidence for recent, population-specific evolution of the human mutation rate. *Proc Natl Acad Sci USA* **112**: 3439-3444
32. Lalancette C, Platts AE, Johnson GD, Emery BR, Carrell DT, Krawetz SA (2009) Identification of human sperm transcripts as candidate markers of male fertility. *J Mol Med* **87**: 735-748
33. Rostkowski M, Olsson MH, Sondergaard CR, Jensen JH (2011) Graphical analysis of pH-dependent properties of proteins predicted using PROPKA. *BMC Struct Biol* **11**: 6
34. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA* **98**: 10037-10041

Figure Legends

Figure 1

A3A/A3B chimeric constructions

A) CLUSTALW alignment of A3A, A3B A3BnA, A3Bc. Mutational clusters are boxed in blue and residues involved in zinc coordination are represented by asterisks. Variable residues are depicted in red. Numbering above is that of A3A while that below is of A3B.

B) Western-blot analysis of V5-tagged APOBEC3 in human HEK 293T cells. β -Actin was used as loading control.

C) Cellular localization of A3 proteins. Confocal microscopy of V5 tagged A3 proteins performed in HeLa cells 24 hours post transfection. Nuclei are stained using DAPI. N: nuclear localization, N/C: nucleocytoplasmic localization.

D) In vitro deamination assay performed on TAMRA–FAM-coupled oligonucleotide using transfected 293T lysates from two experiments performed in duplicates. Background fluorescence obtained with mock-transfected cells was subtracted. RFU, Relative Fluorescence Unit.

E) 3D-PCR gel gradients for HBV DNA editing by A3 proteins.

F) TP53 specific 3DPCR gels after 293T-UGI transfections with A3 proteins.

Figure 2

Schematic representation of A3BnA and A3Bc and cellular localization of A3 proteins

A) Schematic representation of A3BnA and A3Bc derived chimeras.

B) Cellular localization of A3 proteins. Confocal microscopy of V5 tagged A3 proteins performed in HeLa cells 24 hours post transfection. Nuclei are stained using DAPI. N: nuclear localization, N/C: nucleocytoplasmic localization.

C) Fluorescence-activated cell sorting (FACS) analysis of γ H2AX-positive HeLa cells gated on V5-positive cells after A3 48 h post transfection. Error bars represent s.d. from six independent transfections. Differences compared with A3AC106S catalytic mutant were calculated using the Mann–Whitney test (* $P < 0.1$, ** $P < 0.01$, *** $P < 0.001$).

Figure 3

FACS analysis of γ H2AX-positive HeLa cells and representation of the A3A structure

A) Fluorescence-activated cell sorting (FACS) analysis of γ H2AX-positive HeLa cells gated on V5-positive cells after A3B mutants at 48 h post transfection. Error bars represent s.d.

from six independent transfections. Differences compared with A3AC106S catalytic mutant were calculated using the Mann–Whitney test (* $P < 0.1$, ** $P < 0.01$, *** $P < 0.001$).

B) FACS analysis for quadruple and single residue A3Bc mutants. Annotated as in A.

C) FACS analysis for A3Ap2 mutants at 48 h post transfection. Annotated as in A.

D) Cartoon representation of the A3A structure. Side chains of residues identified by mutagenesis are shown as ball-and-sticks. Residues reported to undergo chemical shift upon nucleotide addition are colored in red (>0.05 p.p.m) and yellow (0.028-0.0050 p.p.m). The zinc atom is represented as blue sphere and labeling refers to A3A/A3B numbering.

E) Molecular surface representation of the A3 enzymes, colored according to its electrostatic potential (-5 kT/e [red] to 5 kT/e [blue]) calculated using APBS. Amino acid differences between A3 enzymes have an impact on the electrostatic surface potential (ESP).

Figure 4

Alignment of A3A/A3Bc proteins and cellular localization of Rhesus-A3 proteins

A) CLUSTALW alignment of A3A/A3Bc proteins from human and rhesus monkey. Critical functional residues distinguishing human A3A/A3Bc are boxed in orange and while residues involved in zinc coordination are represented by an asterisk. Numbering above is that of human A3A while that below is of rhA3B.

B) Western-blot analysis of V5-tagged rhesus-A3 proteins in human HEK 293T cells. β -Actin was used as loading control.

C) Cellular localization of Rhesus-A3 proteins. Confocal microscopy of V5 tagged A3 proteins performed in HeLa cells 24 hours post transfection. Nuclei are stained using DAPI.

D) Fluorescence-activated cell sorting (FACS) analysis of γ H2AX-positive HeLa cells gated on V5-positive cells after A3B mutants transfection at 48 h. Error bars represent s.d. from four independent transfections. Differences compared with A3AC106S catalytic mutant were calculated using the Mann–Whitney test (** $P < 0.01$, *** $P < 0.001$).

Expanded View Figure and Table Legends

Figure EV1

Fluorescence-activated cell sorting analysis of γ H2AX-positive cells and western blot analysis of V5-tagged A3 chimeras

A) Fluorescence-activated cell sorting (FACS) analysis of γ H2AX-positive HeLa cells gated on V5-positive cells after A3B mutants transfection at 48 h. Error bars represent s.d. from four independent transfections. Differences compared with A3AC106S catalytic mutant were calculated using the Mann–Whitney test (**P<0.01).

B) Western blot analysis of V5-tagged A3 chimeras in human HEK 293T cells. β -Actin was used as loading control.

Figure EV2

Evolutionary history of A3A and A3B Z1 domain and theoretical Isoelectric point on primate A3 domains

A) Numbers correspond to bootstrap values inferred from 1,000 replicates. Evolutionary distances were computed using Poisson correction method and are expressed in numbers of substitution per site.

B) Theoretical Isoelectric point (pI) performed on primate A3 domains using primary amino acid sequences. Obtained values are rounded-off to one digit after decimal. A3 domains are organized following human A3 domains phylogeny.

Figure EV3

Human polynucleotide cytidine deaminase transcriptome in purified mature human sperm. Absolute numbers of RNAseq transcripts identified. *PRM1* is the sperm specific protamine 1 gene while *HPRT1* is known not to be expressed in sperm.

Table EV1

Compendium of primers used to generate A3A/A3B chimeras coding sequences for pcDNA3.1 TOPO TA V5 cloning.

Table EV2

Compendium of primers used for site directed mutagenesis on A3B expression plasmids.

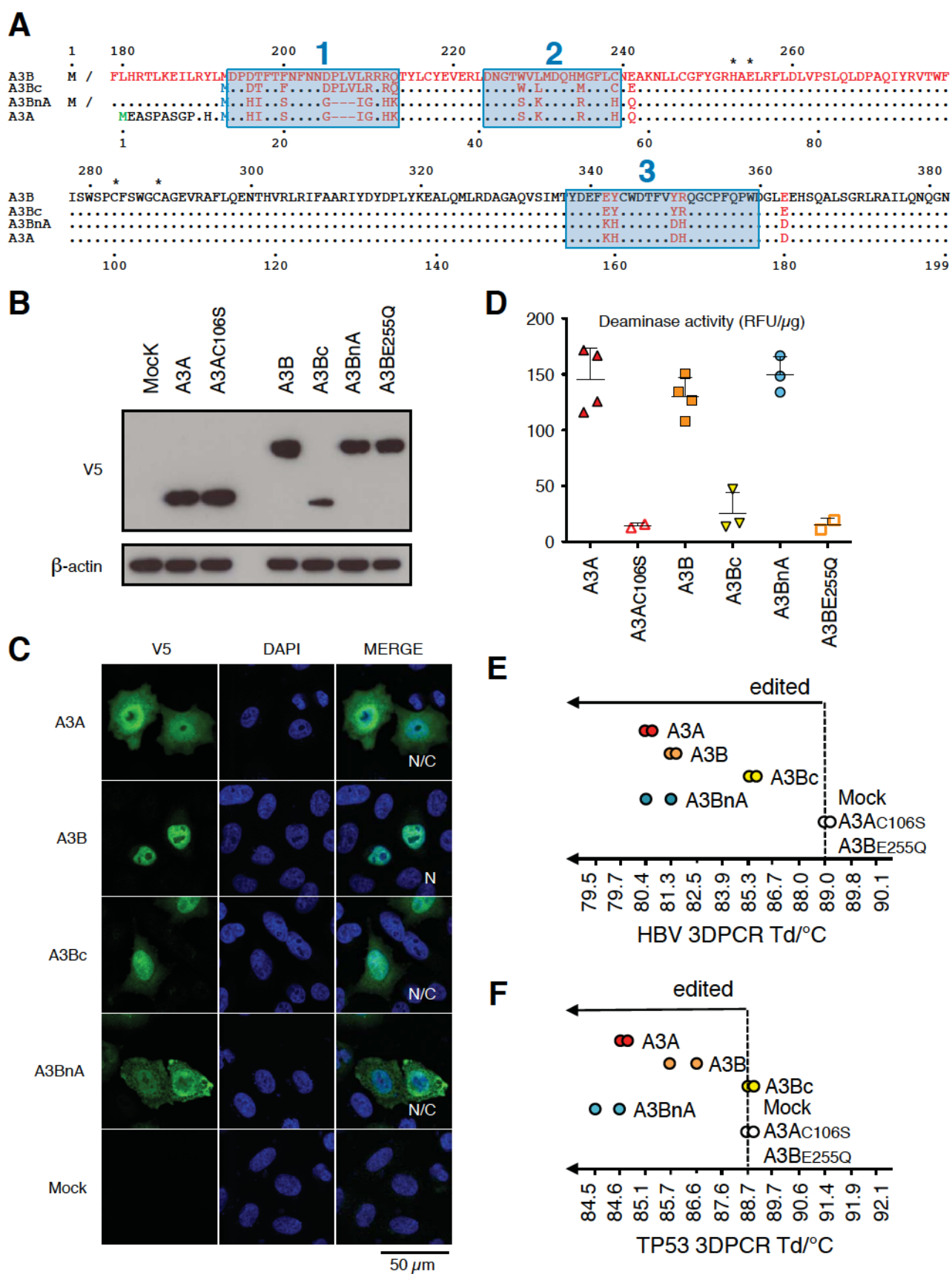
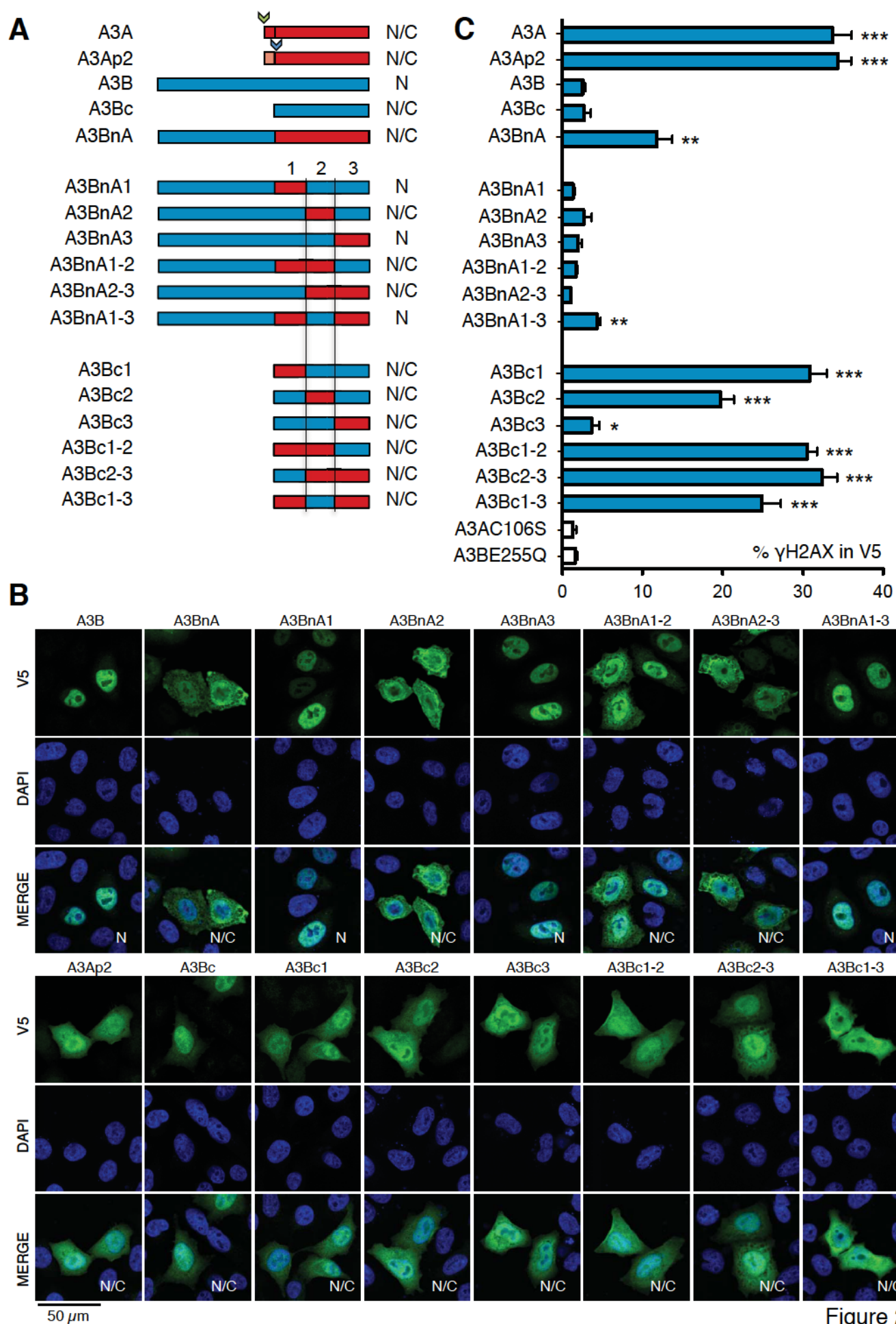


Figure 1



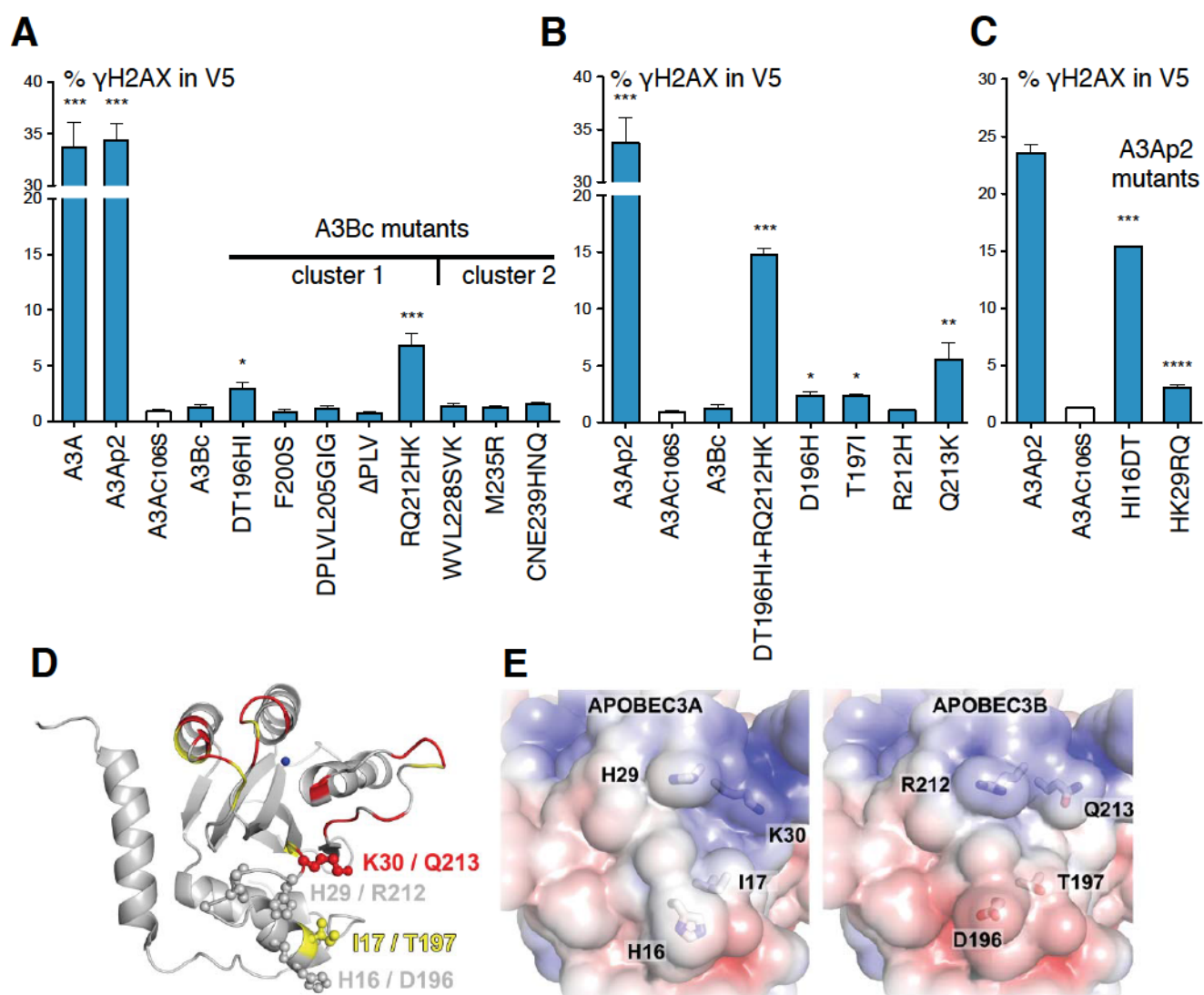


Figure 3

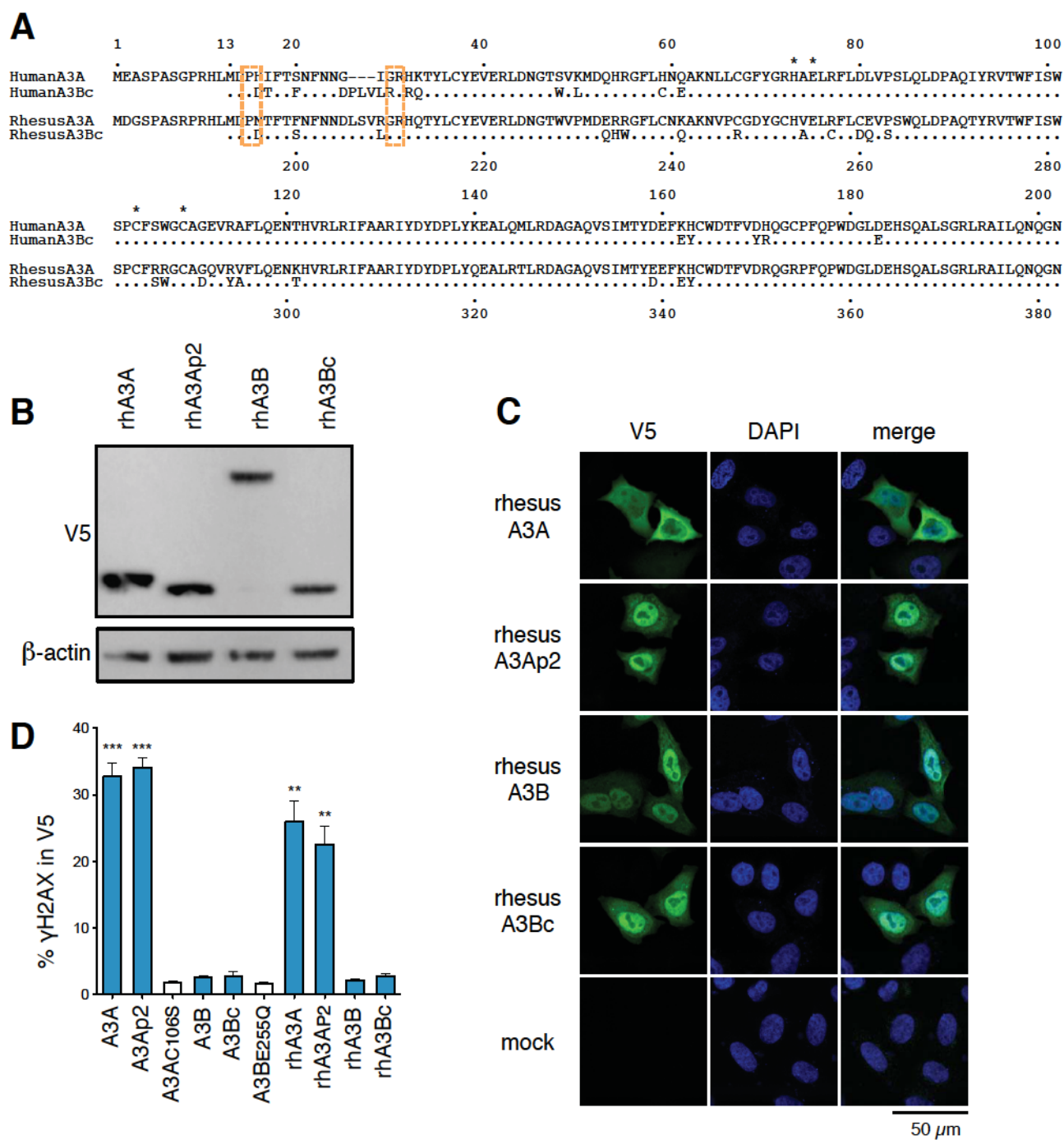


Figure 4

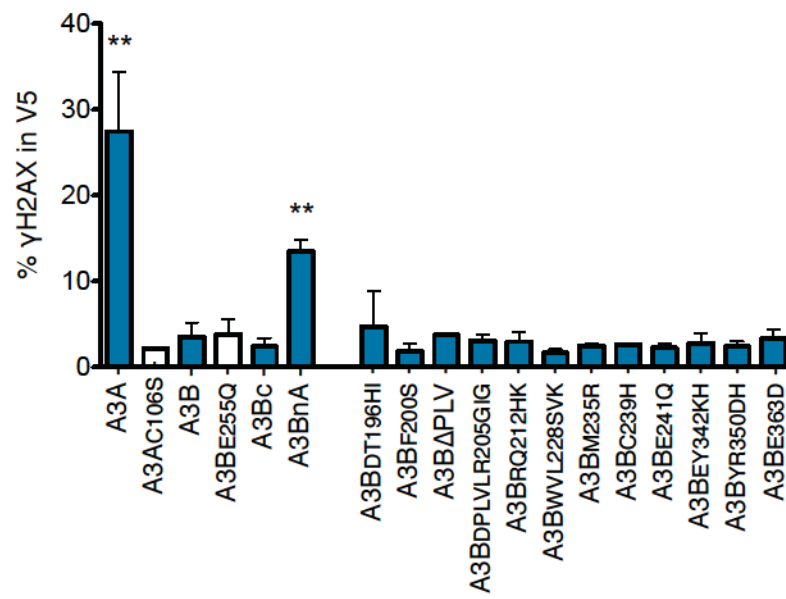
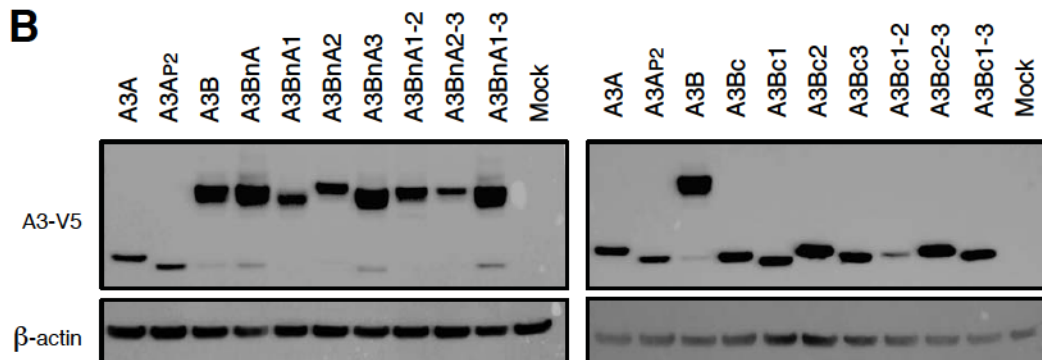
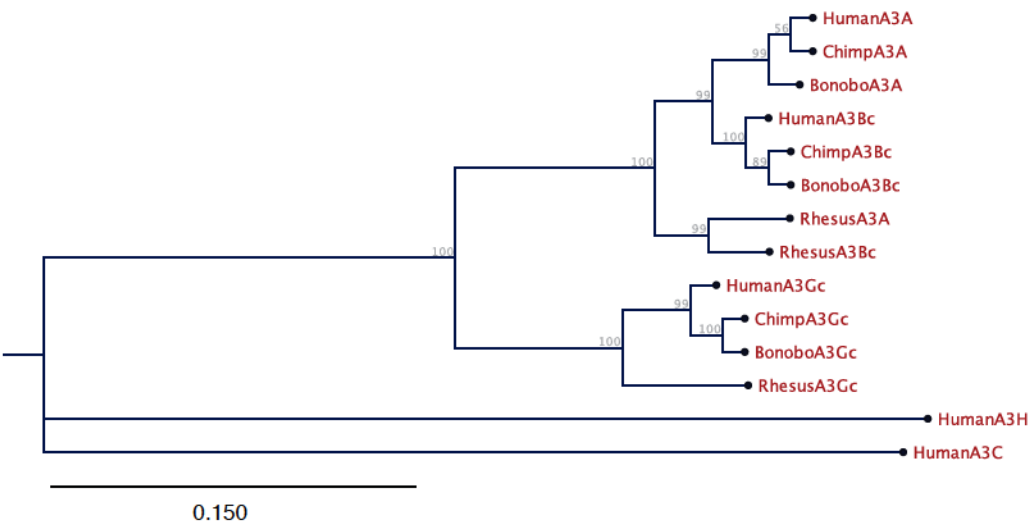
A**B**

Figure EV1

A



B

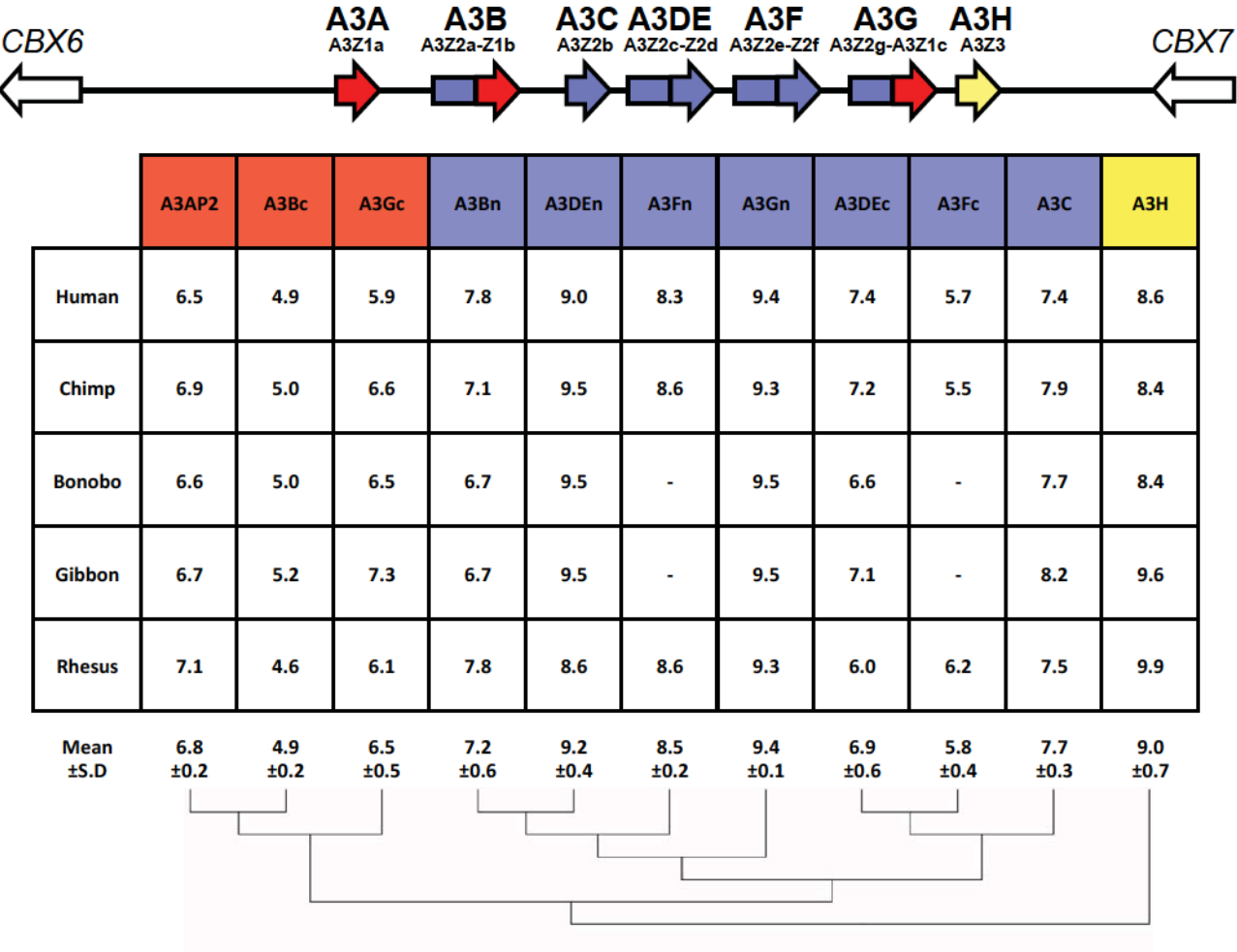


Figure EV2

Plasmid	Matrix	Primers
A3Bc	A3Bi7	For : 5'-CACCATGGATCCAGACACATTCACCTTCAACTTTA-3' Rev : 5'-GTTTCCCTGATTCTGGAGAAT-3'
A3BnA	Puc57-A3BnA	For : 5'-CACCATGAATCCACAGATCAGAAATCCGATGGAGC-3' Rev : 5'-GTTTCCCTGATTCTGGAAGAGCAGGGGGGTG-3'
A3BnA1	A3BnA	For : 5'-CACCATGAATCCACAGATCAGAAATCCG-3' Rev : 5'-GCCATTGTCCAGGCGCTCCACTTCGTAGCACAG-3'
	A3Bi7	For : 5'-CTGTGCTACGAAGTGGAGCGCCTGGACAATGGC-3' Rev : 5'-GTTTCCCTGATTCTGGAAGAGCAGGGGGGTG-3'
A3BnA2	A3BnA	For : 5'-CACCATGAATCCACAGATCAGAAATCCG-3' Rev : 5'-GCCATTGTCCAGGCGCTCCACTTCGTAGCACAG-3'
	A3Bi7	For : 5'-CTGTGCTACGAAGTGGAGCGCCTGGACAATGGC-3' Rev : 5'-TATAAGGAGGCGCTGCAAAATGCTGCGGGATGCT-3'
A3BnA3	A3Bi7	For : 5'-CACCATGAATCCACAGATCAGAAATCCG-3' Rev : 5'-AGCATCCCGCAGCATTTGCAGCGCCTCCTTATA-3'
	A3BnA	For : 5'-TATAAGGAGGCGCTGCAAAATGCTGCGGGATGCT-3' Rev : 5'-GTTTCCCTGATTCTGGAAGAGCAGGGGGGTG-3'
A3BnA1-2	A3BnA	For : 5'-CACCATGAATCCACAGATCAGAAATCCG-3' Rev : 5'-AGCATCCCGCAGCATTTGCAGCGCCTCCTTATA-3'
	A3Bi7	For : 5'-TATAAGGAGGCGCTGCAAAATGCTGCGGGATGCT-3' Rev : 5'-GTTTCCCTGATTCTGGAAGAGCAGGGGGGTG-3'
A3BnA1-3	A3BnA	For : 5'-CACCATGAATCCACAGATCAGAAATCCG-3' Rev : 5'-AGCATCCCGCAGCATTTGCAGCGCCTCCTTATA-3'
	A3Bi7	For : 5'-TATAAGGAGGCGCTGCAAAATGCTGCGGGATGCT-3' Rev : 5'-GTTTCCCTGATTCTGGAAGAGCAGGGGGGTG-3'
A3BnA2-3	A3Bi7	For : 5'-CACCATGAATCCACAGATCAGAAATCCG-3' Rev : 5'-GCCATTGTCCAGGCGCTCCACTTCGTAGCACAG-3'
	A3BnA	For : 5'-CTGTGCTACGAAGTGGAGCGCCTGGACAATGGC-3' Rev : 5'-GTTTCCCTGATTCTGGAAGAGCAGGGGGGTG-3'
A3Bc1	A3BnA1	For : 5'-CACCATGGATCCACACATATTCACCTCC-3'
A3Bc1-2	A3BnA1-2	Rev : 5'-GTTTCCCTGATTCTGGAAGAGCAGGGGGGTG-3'
A3Bc1-3	A3BnA1-3	
A3Bc2	A3BnA2	For : 5'-CACCATGGATCCACACATATTCACCTCC-3'
A3Bc2-3	A3BnA2-3	Rev : 5'-GTTTCCCTGATTCTGGAAGAGCAGGGGGGTG-3'
A3Bc3	A3BnA3	
Rhesus-A3B	Puc57-RhesusA3B	For : 5'-CACCATGAATCCACAGATCAGAAATCCG-3' Rev : 5'-GTTTCCCTGATTCTGGAGAATGGC-3'
Rhesus-A3Bc	Puc57-RhesusA3B	For : 5'-CACCATGGATCCAGACACGTTCACTTCCAA-3' Rev : 5'-GTTTCCCTGATTCTGGAAGAGCAGGGGGGTG-3'
Rhesus-A3Ap2	Rhesus-A3A	For : 5'-CACCATGGATCCAAACAGTTCACTTCAAC-3' Rev : 5'-GTTTCCCTGATTCTGGAGAATGGC-3'
A3Ap2HI16DT	A3A	For : 5'-CACCATGGATCCAGACACATTCACCTT-3' Rev : 5'-GTTTCCCTGATTCTGGAGAATGGC-3'
A3BcD96H	A3Bc	For : 5'-CACCATGGATCCACACACATTCACCTT-3' Rev : 5'-GTTTCCCTGATTCTGGAGAAT-3'
A3BcT97I	A3Bc	For : 5'-CACCATGGATCCAGACATATTCACCTT-3' Rev : 5'-GTTTCCCTGATTCTGGAGAAT-3'

Table EV1

Plasmid	Matrix	Primers
A3BDT196HI	A3Bi7	For : 5'-TACCTGATGGATCCACACATATTCACCTTTCAACTTT-3' Rev : 5'-AAAGTTGAAAGTGAATATGTGTGGATCCATCAGGTA-3'
A3BF200S	A3Bi7	For : 5'-CCAGACACATTCACCTCCAACCTTTAATAATGAC-3' Rev : 5'-GTCATTATTAAAGTTGGAAGTGAATGTGTCTGG-3'
A3BAPLV	A3Bi7	For : 5'-TTCAACTTTAATAATGACCTTCGACGGCGCCAGACC-3' Rev : 5'-GGTCTGGCGCCGTCGAAGGTCATTATTAAAGTTGAA-3'
A3BDPLVLR205GIG	A3Bi7	For : 5'-TTCAACTTTAATAATGGCATTGGACGGCGCCAGACCTAC-3' Rev : 5'-GTAGGTCTGGCGCCGTCGAATGCCATTATTAAAGTTGAA-3'
A3BRQ212HK	A3Bi7 A3BDT196HI	For : 5'-TTGGTCCTTCGACGGCATAAGACCTACTTGTGCTAT-3' Rev : 5'-ATAGCACAAAGTAGGTCTTATGCCGTCGAAGGACCAA-3'
A3BWVL228SVK	A3Bi7	For : 5'-CTGGACAATGGCACCTCGGTCAAGATGGACCAGCACATG-3' Rev : 5'-CATGTGCTGGTCCATCTTGACCGAGGTGCCATTGTCCAG-3'
A3BM235R	A3Bi7	For : 5'-CTGATGGACCAGCACAGGGGCTTTCTATGCAAC-3' Rev : 5'-GTTGCATAGAAAGCCCTGTGCTGGTCCATCAG-3'
A3BC239H	A3Bi7	For : 5'-CACATGGGCTTTCTACACAACGAGGCTAAGAAT-3' Rev : 5'-ATTCTTAGCCTCGTTGTGTAGAAAGCCCATGTG-3'
A3BE241Q	A3Bi7	For : 5'-GGCTTTCTATGCAACCAGGCTAAGAATCTTCTCTGT-3' Rev : 5'-ACAGAGAAGATTCTTAGCCTGGTTGCATAGAAAGCC-3'
A3BCNE239HNQ	A3Bi7	For : 5'-CACATGGGCTTTCTACACAACGAGGCTAAGAATCTTCTCTGT-3' Rev : 5'-ACAGAGAAGATTCTTAGCCTGGTTGTGTAGAAAGCCCATGTG-3'
A3BEY342KH	A3Bi7	For : 5'-ACCTACGATGAGTTTAAGCACTGCTGGGACACCTTT-3' Rev : 5'-AAAGGTGTCCAGCAGTGCTTAAACTCATCGTAGGT-3'
A3BYR350DH	A3Bi7	For : 5'-TGGGACACCTTTGTGGACCACCAGGGATGTCCCTTC-3' Rev : 5'-GAAGGGACATCCCTGGTGGTCCACAAAGGTGTCCCA-3'
A3BE363D	A3Bi7	For : 5'-CCCTGGGATGGACTAGATGAGCACAGCCAAGCC-3' Rev : 5'-GGCTTGGCTGTGCTCATCTAGTCCATCCCAGGG-3'
A3BcR212H	A3Bc	For : 5'-TTGGTCCTTCGACGGCATCAGACCTACTTGTGCTAT-3' Rev : 5'-ATAGCACAAAGTAGGTCTGATGCCGTCGAAGGACCAA-3'
A3BcQ213K	A3Bc	For : 5'-TTGGTCCTTCGACGGCGCAAGACCTACTTGTGCTAT-3' Rev : 5'-ATAGCACAAAGTAGGTCTTGCGCCGTCGAAGGACCAA-3'

Table EV2

Keywords: APOBEC3; cytidine deaminase; hypermutation; cancer genomics; 3DPCR

Erroneous identification of APOBEC3-edited chromosomal DNA in cancer genomics

R Suspène¹, V Caval¹, M Henry¹, M S Bouzidi¹, S Wain-Hobson¹ and J-P Vartanian^{*,1}

¹Molecular Retrovirology Unit, Institut Pasteur, 28 rue du Dr. Roux, 75724 Paris, France

Background: The revolution in cancer genomics shows that the dominant mutations are CG->TA transitions. The sources of these mutations are probably two host cell cytidine deaminases APOBEC3A and APOBEC3B. The former in particular can access nuclear DNA and monotonously introduce phenomenal numbers of C->T mutations in the signature 5'TpC context. These can be copied as G->A transitions in the 5'GpA context.

Methods: DNA hypermutated by an APOBEC3 enzyme can be recovered by a technique called 3DPCR, which stands for differential DNA denaturation PCR. This method exploits the fact that APOBEC3-edited DNA is richer in A + T compared with the reference. We explore explicitly 3DPCR error using cloned DNA.

Results: Here we show that the technique has a higher error rate compared with standard PCR and can generate DNA strands containing both C->T and G->A mutations in a 5'GpCpR context. Sequences with similar traits have been recovered from human tumour DNA using 3DPCR.

Conclusions: Differential DNA denaturation PCR cannot be used to identify fixed C->T transitions in cancer genomes. Presently, the overall mutation frequency is $\sim 10^4$ – 10^5 base substitutions per cancer genome, or 0.003 – 0.03 kb^{-1} . By contrast, the 3DPCR error rate is of the order of 4 – 20 kb^{-1} owing to constant selection for AT DNA and PCR-mediated recombination. Accordingly, sequences recovered by 3DPCR harbouring mixed C->T and G->A mutations associated with the 5'GpC represent artefacts.

The advent of cancer genomics has shown that there are far more somatic mutations than hitherto imagined. Although large karyotypic changes have been known for decades, there are a myriad of small rearrangements, whereas the number of mutations can exceed 10^5 per genome (Greenman *et al*, 2007; Stephens *et al*, 2009, 2012; Pleasance *et al*, 2010). Apart from cancers involving mutagens that leave genetic signatures, such as aflatoxin or benzo[a]pyrene, overall there is an excess of CG->TA transitions in cancer genomes (Greenman *et al*, 2007). For some genomes there is a pronounced 5' base context, notably TpC for C->T transitions (Nik-Zainal *et al*, 2012; Stephens *et al*, 2012; Alexandrov *et al*, 2013). Of course, the G->A transition is merely a C->T transition on the opposite strand, suggesting a single transition.

The human APOBEC3 (A3) family of seven genes encoding six functional cytidine deaminases (APOBEC3A, A3B, A3C, A3F, A3G and A3H) has come to the fore in recent years as host cell mutators

of viral DNA (Jarmuz *et al*, 2002; Sheehy *et al*, 2002; Harris *et al*, 2003; Lecossier *et al*, 2003; Mangeat *et al*, 2003; Mariani *et al*, 2003). They are closely related to activation-induced deaminase (AID), which is responsible for immunoglobulin class switch recombination and somatic hypermutation or rearranged VDJ genes (Di Noia and Neuberger, 2007). Indeed, it is thought that the A3 locus emerged by gene duplication of AID (Conticello *et al*, 2005). Like AID, the substrate specificity of the A3 enzymes is the cytidine base in single-stranded DNA (ssDNA), the product being uridine, which base pairs as thymidine. As such, they effectively hypermutate DNA, the mutational load being so great that the viral genetic information is irretrievably lost. That several A3 genes are upregulated by type I and II interferons lent credence to their being novel restriction factors and part of a broad innate immune response to microbes (Bonvin *et al*, 2006; Koning *et al*, 2009; Refsland *et al*, 2010; Stenglein *et al*, 2010). These APOBEC3 mutator enzymes must be tightly controlled as at least four (A3A,

*Correspondence: Dr J-P Vartanian; E-mail: jean-pierre.vartanian@pasteur.fr

Received 23 October 2013; revised 26 February 2014; accepted 10 March 2014; published online 1 April 2014

© 2014 Cancer Research UK. All rights reserved 0007–0920/14

A3B, A3C and A3H) can access the nucleus (Vartanian *et al*, 2008; Stenglein *et al*, 2010; Landry *et al*, 2011; Suspène *et al*, 2011a; Aynaud *et al*, 2012; Shinohara *et al*, 2012; Burns *et al*, 2013a, b).

More recently, it has emerged that APOBEC3A (A3A) and probably APOBEC3B (A3B) can mutate nuclear DNA (nuDNA) (Suspène *et al*, 2011a; Aynaud *et al*, 2012; Shinohara *et al*, 2012; Burns *et al*, 2013a). A3A can edit both cytidine and 5-methylcytidine residues in ssDNA (Carpenter *et al*, 2012; Wijesinghe and Bhagwat, 2012; Suspène *et al*, 2013) and can generate DNA double-strand breaks (Landry *et al*, 2011; Mussil *et al*, 2013). Low levels of A3 mutation that do not overwhelm DNA mismatch repair are probably the source of the CG->TA mutations in cancer genomes. By contrast, hypermutation is proapoptotic and represents a novel pathway for DNA catabolism (Suspène *et al*, 2011a; Mussil *et al*, 2013). Hypermutated nuDNA with hundreds of mutations per kilobase can be found in both pathological and physiological settings (Suspène *et al*, 2011a). Invariably the mutations are influenced by the base 5' with a TpC and CpC being the preferred targets (the edited cytidine residue being underlined) (Suspène *et al*, 2011a; Nik-Zainal *et al*, 2012; Alexandrov *et al*, 2013). Only A3G shows a strong penchant for CpC (Beale *et al*, 2004; Bishop *et al*, 2004; Suspène *et al*, 2004; Henry *et al*, 2009; Vartanian *et al*, 2010). By contrast, AID prefers a purine residue in the 5' position, where GpC>ApC (Conticello *et al*, 2005; Vartanian *et al*, 2010). At low levels of nuDNA editing, compatible with DNA repair, these enzymes are probably the source of the C->T transitions seen in cancer genomes (Greenman *et al*, 2007; Stephens *et al*, 2009, 2012; Pleasance *et al*, 2010; Nik-Zainal *et al*, 2012). While the relative contribution of A3A and A3B to the editing of nuDNA needs to be ascertained, the A3B^{-/-} genotype is particularly prevalent in SE Asia (Kidd *et al*, 2007). Interestingly, these individuals have a higher odds ratio of developing breast and liver cancer compared with controls (Komatsu *et al*, 2008; Zhang *et al*, 2012).

DNA that has been hypermutated by an APOBEC3 (A3) enzyme can be recovered by a technique called 3DPCR, which stands for differential DNA denaturation PCR (Suspène *et al*, 2005a). This method exploits the fact that A3-edited DNA is richer in A+T compared with the reference. Modulation of the PCR denaturation temperature allows selective amplification of AT-rich DNA, sometimes by up to 10⁴-fold (Suspène *et al*, 2005b).

While 3DPCR was first used to detect hypermutated viral genomes (Suspène *et al*, 2005b; Vartanian *et al*, 2008, 2010; Suspène *et al*, 2011b), it is being increasingly used to recover low levels of A3-edited nuclear DNA. Two recent papers are noteworthy because they used 3DPCR to recover lightly edited DNA sequences manifesting ~2–13 CG->TA transitions per kilobase of nuDNA (Shinohara *et al*, 2012; Burns *et al*, 2013a). They exploit the use of 4–5 °C temperature gradients across the 12-wells of the heating block, which is small in comparison with the ±0.4 °C temperature stability per well leaving considerable room for experimental variation. One study presented the sequences in sufficient detail – the sequences were not monotonously substituted by C->T or G->A (Shinohara *et al*, 2012); rather they contained both mutations on the same strand. Furthermore, a preferred 5'GpC editing context was frequently noted, which is more typical of AID rather than an A3 enzyme (Beale *et al*, 2004). Although ectopic upregulation of AID has indeed been reported (Endo *et al*, 2007; Matsumoto *et al*, 2007; Morisawa *et al*, 2008), we were concerned that such low-editing rates per kilobase might be close to the 3DPCR error threshold. The second study implicated A3B in nuDNA editing (Burns *et al*, 2013a), yet the signal was only 2–3-fold over 3DPCR background, which we found surprisingly low, particularly as A3B can extensively edit HBV DNA, the editing rates being of the order of >100 kb⁻¹ (Suspène *et al*, 2005b; Vartanian *et al*, 2010).

Here we explore explicitly 3DPCR error using cloned DNA. It transpires that the 3DPCR error rate is of the order of 4–20 kb⁻¹ and generates sequences encoding both C->T and G->A substitutions occurring preferentially in the GpC dinucleotide. Sequences with similar traits have been recovered from human DNA (Shinohara *et al*, 2012; Burns *et al*, 2013a).

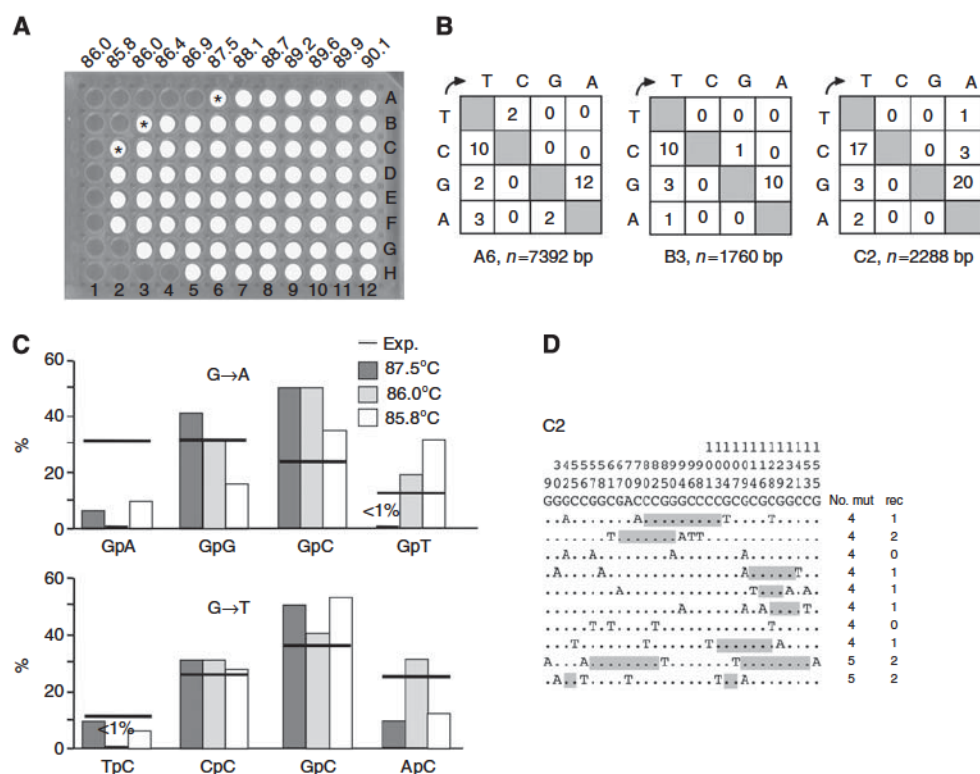
MATERIALS AND METHODS

Blood was obtained from anonymous healthy donors and peripheral blood mononuclear cells (PBMCs) were isolated by Ficoll gradient (Eurobio AbCys, Courtaboeuf, France). Isolation of CD4⁺ T lymphocytes was performed by incubation with antibody-coated magnetic beads (Miltenyi Biotec, Bergish Gladbach, Germany). Purity of CD4⁺ T lymphocytes was above 90% as checked by flow cytometry (FACSCalibur; Becton Dickinson, Franklin Lakes, NJ, USA). CD4⁺ T lymphocytes were stimulated with 10 µg ml⁻¹ PHA (Sigma, St Louis, MO, USA), 100 U ml⁻¹ IL-2 (Sigma) and 500 U ml⁻¹ IFN-α (PBL Biomedical Laboratories, Piscataway, NJ, USA) for 48 h. For the detection of hypermutation by 3DPCR, CD4⁺ T cells were transduced with lentivirus rV2.EF1.UGI, which encodes a codon optimised UNG inhibitor (UGI) under the control of the constitutive human EF1 promoter (Vectalys, Toulouse, France).

A molecular clone corresponding to 376 bp spanning exon 8 and intron 9 of the human TP53 gene was recovered from HeLa cells. For amplification of human TP53, input DNA was ~1 ng. The first-round reaction parameters were 95 °C for 5 min, followed by 40 cycles (95 °C for 30 s, 60 °C for 30 s and 72 °C for 2 min), and finally 10 min at 72 °C with the following primers: P53ext5 (5'-GAGCTGGACCTTAGGCTCAGAAAGGACAA-3') and P53ext3 (5'-GCTGGTGTGTTGGGCA GTGCTAGGAA-3'). Second-round 3DPCR was performed with 1/50 of the first round as input using an Eppendorf Mastercycler ep Gradient S (Eppendorf AG, Hamburg, Germany) programmed to generate a 4 °C gradient in the denaturation temperature. The reaction parameters were 86–90 °C for 5 min, followed by 42 cycles (86–90 °C for 30 s, 58 °C for 30 s and 72 °C for 2 min), and finally 10 min at 72 °C. The buffer conditions were 2.5 mM MgCl₂, 16 mM (NH₄)₂SO₄, 67 mM Tris-HCl (pH 8.8), 0.01% Tween-20, 200 µM each deoxynucleoside triphosphate, 100 µM each primer (P53int5, 5'-TTCTCTTTTCTAT CCTGAGTAGTGGTAA-3'; P53int3, 5'-AAAGGTGATAAAAGTG AATCTGAGGCATAA-3') and 1.5 U of Taq (EurobioTaq +; Eurobio AbCys) or Pfu (PfuUltra II Fusion HS DNA Polymerase; Agilent Technologies, Santa Clara, CA, USA) DNA polymerase.

Conditions of amplification of the β-catenin were 95 °C for 5 min, followed by 35 cycles (95 °C for 30 s, 60 °C for 30 s and 72 °C for 10 min) and finally 20 min at 72 °C with the following primers: 5'βcatout (5'-AGCTGATTTGATGGAGTTGGACA-3') and 3'βcatout (5'-CCAGCTACTTGTCTTGTAGTGAA-3'). Nested PCR was performed with 1/50 of the first round, conditions were 80–87 °C for 5 min, followed by 35 cycles (80–87 °C for 30 s, 60 °C for 30 s and 72 °C for 10 min) and finally 20 min at 72 °C with the following primers: 5'βcatin (5'-ACATGGCCATGGAAC CAGACAGA-3') and 3'βcatin (5'-GTTCTTGTAGTGAAGGACT GAGAA-3').

For long-range PCR, the first-round reaction parameters were 95 °C for 5 min, followed by 40 cycles (95 °C for 30 s, 60 °C for 30 s and 72 °C for 10 min), and finally 20 min at 72 °C. The reaction parameters for the second-round 3DPCR were 86–90 °C for 5 min, followed by 42 cycles (86–90 °C for 30 s, 58 °C for 30 s and 72 °C for 10 min), and finally 20 min at 72 °C. The buffer conditions were 2.5 mM MgCl₂, 1 × LA PCR Buffer II (TaKaRa LA Taq; TaKaRa, Otsu, Japan), 400 µM each deoxynucleoside triphosphate (TaKaRa LA Taq; TaKaRa), 100 µM each primer and 1.5 U of Taq (EurobioTaq +; Eurobio AbCys).



performed using a 7 °C denaturation gradient (80–87 °C) and *Taq* polymerase. DNA from the last positive amplification was cloned and sequenced.

Once again sequences with both C->T and G->A substitutions dominated with up to four such substitutions per sequence, although there were a few sequences encoding just one type of transition (Figure 3A and B). The overall mutation frequency for sample no. 310 was $\sim 3 \times$ that for sample nos. 326, 345 and HL2 and more than 13 x the plasmid control (Figure 3C). Once again the GpC context predominated (Figure 3D), which argues against the involvement of an A3 enzyme.

Is it possible to distinguish lightly A3-edited sequences from the 3DPCR background described above, and what would an unambiguous A3-edited β -catenin look like? As haematopoietic cells generally show high A3 expression levels compared with other tissues (<http://www.biogps.org>), we analysed DNA from purified CD4⁺ T cells from PBMCs that had been treated by PHA, IL-2 and interferon- α , which strongly upregulates the A3A gene (Aynaud *et al*, 2012). To increase detection of A3-edited nuDNA, the cells were transduced with a lentivirus encoding the uracil N-glycosylase inhibitor UGI to block catabolism of dU-containing DNA (Wang and Mosbaugh, 1988). Using a 7°C gradient 3DPCR amplification of the same β -catenin gene segment revealed a signal differing by 1°C with respect to the non-stimulated CD4⁺ T cells as a control (Figure 4A and C). Cloning and sequencing revealed monotonously C->T or G->A substituted sequences (Figure 4A–C) with editing focused on T_{pC} (G_{pA}) dinucleotides, all consistent with editing by A3A and/or A3B (Figure 4E and F).

The non-stimulated control proved interesting. The last positive 3DPCR amplification was at 83.7 °C in the centre of the block to minimise the effects of thermal heterogeneity (Figure 4A and B). A collection of sequences was found that could be broken down into two groups of AT-rich sequences (Figure 4C). The first group of 10 sequences corresponded to what we can call 3DPCR background sequences containing up to five substitutions per locus (32 kb^{-1}) involving both C->T and G->A transitions, there being a bias for transitions in GpC dinucleotides (not shown). The second group of two sequences in Figure 4C were monotonously G->A substituted compared with the plus strand with five and six transitions, most of which were in TpC and CpC. Indeed, the dinucleotide context is very similar to that of the *bona fide* A3-edited β -catenin sequences recovered from IFN- α -stimulated CD4⁺ T cells (Figure 4A). Accordingly, it appears that 3DPCR is indeed capable of recovering lightly A3-edited sequences when present.

The constant presence of sequences bearing both C->T and G->A transitions was intriguing, no matter what the source of input DNA was - molecularly cloned, and tissue- or PBMC-derived DNA. We have previously shown that recombination during PCR can contribute to the complexity of amplified DNA (Meyerhans *et al*, 1990). It usually occurs towards the end of amplification when [DNA] > [Taq], the polymerase simply being unable to complete synthesis of the nascent strands before the next temperature shift to denaturation.

Long elongation times reduce recombination but do not eliminate the problem (Meyerhans *et al.*, 1990). In the above

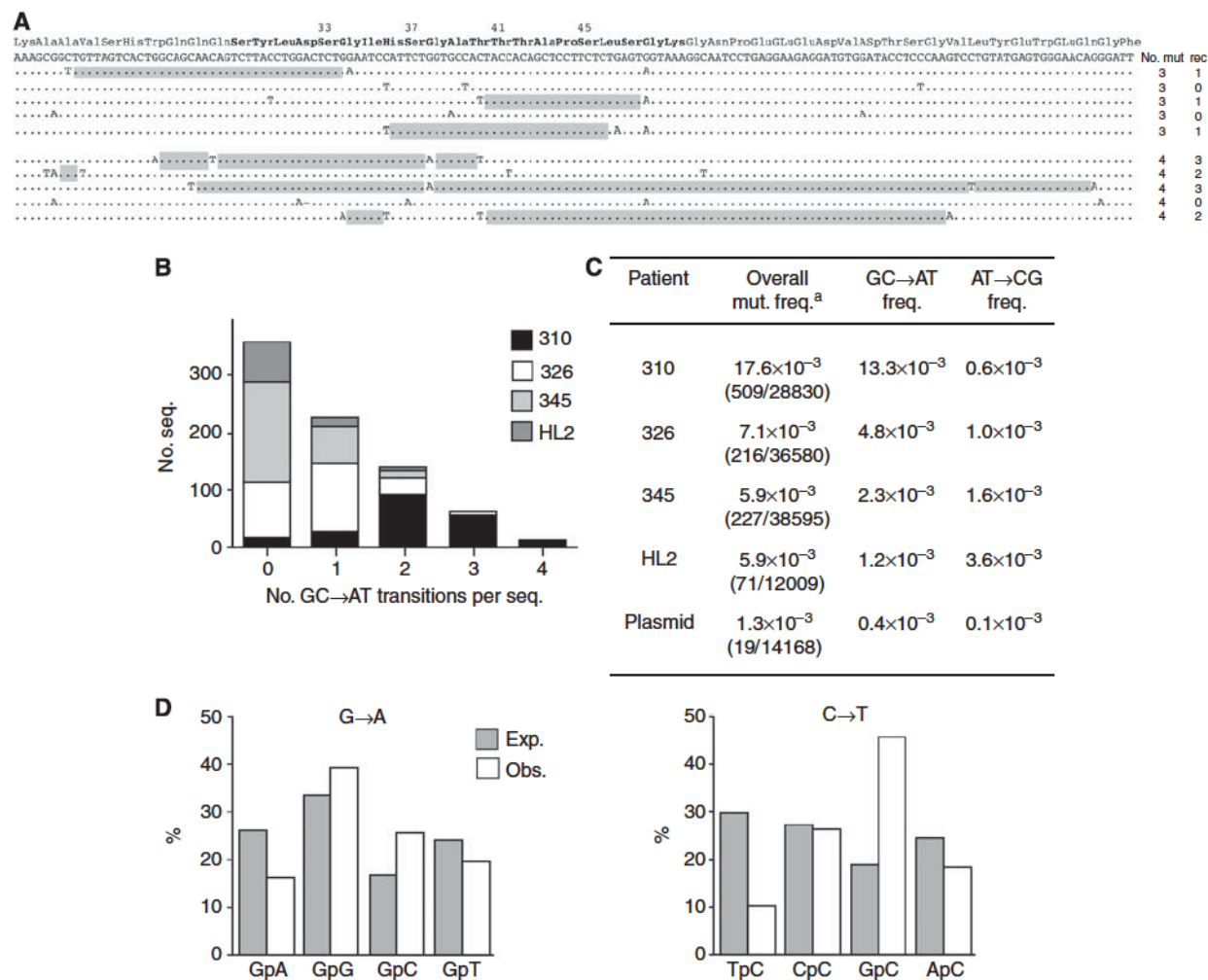


Figure 3. Differential DNA denaturation PCR-derived *CTNNB1* sequences from cirrhotic and normal liver DNA. (A) A collection of 3DPCR recovered *CTNNB1* sequences from viral-associated liver cirrhosis. The amino-acid sequence is shown above where bold case identifies the residues that frequently are substituted in liver cancers. The annotation is as for Figure 1D. (B) Stack-up graph of the number of GC→AT transitions per sequence per liver sample. (C) Mutation frequencies per sample along with the absolute number of mutations and bases sequences (within parentheses). (D) 5' Dinucleotide context of GC→AT transitions for G→A and C→T transitions respectively.

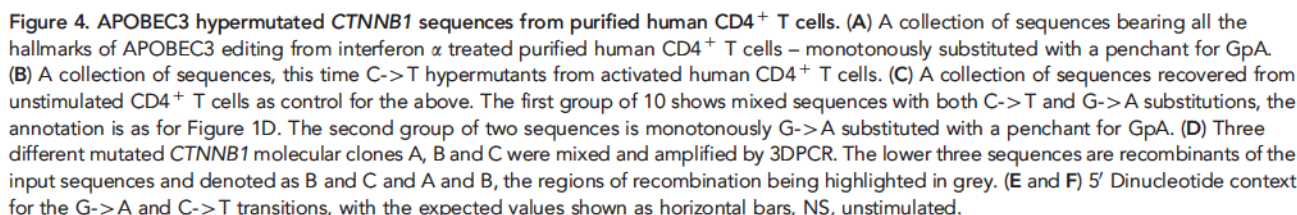
experiments involving *TP53* and *CTNNB1* loci, 10 min elongation times were used. However, as 3DPCR amplification occurs very close to the denaturation temperature, some templates may not be fully denatured, which would block polymerase elongation, thus rendering amplification less efficient. Indeed, this is reflected by the fact that we regularly use 42 cycles when performing 3DPCR, when normally 30 cycles would be more than sufficient to amplify the *TP53* target.

To address the question of PCR recombination, 1 ng of DNA of three different molecularly cloned *CTNNB1* sequences were mixed at an equimolar ratio and 3DPCR performed using the same 7 °C gradient. The last positive amplification (84.7 °C) was cloned and sequenced. By comparison of the input and output sequences (Figure 4D), it is clear that PCR-mediated recombination can increase the sequence complexity. The most parsimonious explanation for hypermutated sequences bearing both C→T and G→A transitions is that thermostable polymerases make a spectrum of mutations. For C→T and G→A transitions, one is dominant and occurs either in the context GpCpR (C→T transition) or YpGpC for a G→A transition. Successive cycles of 3DPCR enrich DNA bearing CG→TA transitions, and to a

lesser extent N→TA transversions are selectively amplified. The complexity of the sequences is increased by PCR recombination.

Differential DNA denaturation PCR was initially used to recover A3 edited viral genomes and plasmid DNA (Vartanian *et al*, 2008, 2010; Suspène *et al*, 2011a, b). Subsequently, it proved crucial to demonstrating that mitochondrial and nuclear DNA could be hyperedited (Suspène *et al*, 2011a). Recently two studies have reported A3 editing of nuDNA following recovery by 3DPCR (Shinohara *et al*, 2012; Burns *et al*, 2013a). The overall mutation rates were within the 3DPCR error range reported here. For one study the difference in the 3DPCR temperature was never > 1 °C compared to the control while the mutation frequencies were 6–9 kb⁻¹, only 2 fold greater than the basal error rate (Shinohara *et al*, 2012). Sequences bearing mixed C→T and G→A substitutions were found while the preferred 5' dinucleotide context for C→T substitutions was GpC, observations reminiscent of the data shown in Figures 1–3 and suggests that they are non-physiological 3DPCR generated sequences.

If mixed DNA molecules resulted from A3 editing then the first edited strand would have to be replicated, followed by strand separation and APOBEC3 editing. The crucial difference compared



Differential DNA denaturation PCR is very powerful at selectively amplifying A3 hyperedited DNA, but as shown above there is a grey zone where the editing frequency falls below 5–15 substitutions per kb⁻¹. Having worked with it for many years (Suspène *et al.*, 2005a, 2011a; Vartanian *et al.*, 2008, 2010), we would suggest that cloning and sequencing is vital to a correct interpretation of the results. *Prima facie* evidence of A3A and A3B

Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR (2013) Deciphering signatures of mutational processes operative in human cancer. *Cell Rep* 3(1): 246–259.

Aynaud MM, Suspene R, Vidalain PO, Mussil B, Guetard D, Tangy F, Wain-Hobson S, Vartanian JP (2012) Human Tribbles 3 protects nuclear DNA from cytidine deamination by APOBEC3A. *J Biol Chem* 287(46): 39182–39192.

- Beale RC, Petersen-Mahrt SK, Watt IN, Harris RS, Rada C, Neuberger MS (2004) Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra *in vivo*. *J Mol Biol* **337**(3): 585–596.
- Bishop KN, Holmes RK, Sheehy AM, Davidson NO, Cho SJ, Malim MH (2004) Cytidine deamination of retroviral DNA by diverse APOBEC proteins. *Curr Biol* **14**(15): 1392–1396.
- Bonvin M, Achermann F, Greeve I, Stroka D, Keogh A, Inderbitzin D, Candinas D, Sommer P, Wain-Hobson S, Vartanian JP, Greeve J (2006) Interferon-inducible expression of APOBEC3 editing enzymes in human hepatocytes and inhibition of hepatitis B virus replication. *Hepatology* **43**(6): 1364–1374.
- Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, Leonard B, Refsland EW, Kotandeniya D, Tretyakova N, Nikas JB, Yee D, Temiz NA, Donohue DE, McDougall RM, Brown WL, Law EK, Harris RS (2013a) APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* **494**(7437): 366–370.
- Burns MB, Temiz NA, Harris RS (2013b) Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat Genet* **45**(9): 977–983.
- Carpenter MA, Li M, Rathore A, Lackey L, Law EK, Land AM, Leonard B, Shandilya SM, Bohn MF, Schiffer CA, Brown WL, Harris RS (2012) Methylcytosine and normal cytosine deamination by the foreign dna restriction enzyme APOBEC3A. *J Biol Chem* **287**(41): 34801–34808.
- Coticello SG, Thomas CJ, Petersen-Mahrt SK, Neuberger MS (2005) Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol Biol Evol* **22**(2): 367–377.
- Datta J, Kutay H, Nasser MW, Nuovo GJ, Wang B, Majumder S, Liu CG, Volinia S, Croce CM, Schmittgen TD, Ghoshal K, Jacob ST (2008) Methylation mediated silencing of microRNA-1 gene and its role in hepatocellular carcinogenesis. *Cancer Res* **68**(13): 5049–5058.
- Di Noia JM, Neuberger MS (2007) Molecular mechanisms of antibody somatic hypermutation. *Annu Rev Biochem* **76**: 1–22.
- Eckert KA, Kunkel TA (1991) DNA polymerase fidelity and the polymerase chain reaction. *PCR Methods Appl* **1**(1): 17–24.
- Eckert KA, Kunkel TA (1993) Effect of reaction pH on the fidelity and processivity of exonuclease-deficient Klenow polymerase. *J Biol Chem* **268**(18): 13462–13471.
- Endo Y, Marusawa H, Kinoshita K, Morisawa T, Sakurai T, Okazaki IM, Watashi K, Shimotohno K, Honjo T, Chiba T (2007) Expression of activation-induced cytidine deaminase in human hepatocytes via NF-kappaB signaling. *Oncogene* **26**(38): 5587–5595.
- Greenman C, Stephens P, Smith R, Dalgleish GL, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens C, Edkins S, O'Meara S, Vastrik I, Schmidt EE, Avis T, Barthorpe S, Bhamra G, Buck G, Choudhury B, Clements J, Cole J, Dicks E, Forbes S, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Jenkinson A, Jones D, Menzies A, Mironenko T, Perry J, Raine K, Richardson D, Shepherd R, Small A, Tofts C, Varian J, Webb T, West S, Widada S, Yates A, Cahill DP, Louis DN, Goldstraw P, Nicholson AG, Brasseur F, Looijenga L, Weber BL, Chiew YE, DeFazio A, Greaves MF, Green AR, Campbell P, Birney E, Easton DF, Chenevix-Trench G, Tan MH, Khoo SK, Teh BT, Yuen ST, Leung SY, Wooster R, Futreal PA, Stratton MR (2007) Patterns of somatic mutation in human cancer genomes. *Nature* **446**(7132): 153–158.
- Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, Neuberger MS, Malim MH (2003) DNA deamination mediates innate immunity to retroviral infection. *Cell* **113**(6): 803–809.
- Henry M, Guetard D, Suspene R, Rusniok C, Wain-Hobson S, Vartanian JP (2009) Genetic editing of HBV DNA by monodomain human APOBEC3 cytidine deaminases and the recombinant nature of APOBEC3G. *PLoS One* **4**(1): e4277.
- Jarmuz A, Chester A, Bayliss J, Gisbourne J, Dunham I, Scott J, Navaratnam N (2002) An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics* **79**(3): 285–296.
- Kidd JM, Newman TL, Tuzun E, Kaul R, Eichler EE (2007) Population stratification of a common APOBEC gene deletion polymorphism. *PLoS Genet* **3**(4): e63.
- Komatsu A, Nagasaki K, Fujimori M, Amano J, Miki Y (2008) Identification of novel deletion polymorphisms in breast cancer. *Int J Oncol* **33**(2): 261–270.
- Koning FA, Newman EN, Kim EY, Kunstman KJ, Wolinsky SM, Malim MH (2009) Defining APOBEC3 expression patterns in human tissues and hematopoietic cell subsets. *J Virol* **83**(18): 9474–9485.
- Landry S, Narvaiza I, Linfesty DC, Weitzman MD (2011) APOBEC3A can activate the DNA damage response and cause cell-cycle arrest. *EMBO Rep* **12**(5): 444–450.
- Lecossier D, Bouchonnet F, Clavel F, Hance AJ (2003) Hypermutation of HIV-1 DNA in the absence of the Vif protein. *Science* **300**(5622): 1112.
- Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D (2003) Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* **424**(6944): 99–103.
- Mariani R, Chen D, Schrofelbauer B, Navarro F, König R, Bollman B, Munk C, Nymark-McMahon H, Landau NR (2003) Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. *Cell* **114**(1): 21–31.
- Matsumoto Y, Marusawa H, Kinoshita K, Endo Y, Kou T, Morisawa T, Azuma T, Okazaki IM, Honjo T, Chiba T (2007) *Helicobacter pylori* infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium. *Nat Med* **13**(4): 470–476.
- Meyerhans A, Vartanian JP, Wain-Hobson S (1990) DNA recombination during PCR. *Nucleic Acids Res* **18**(7): 1687–1691.
- Miyoshi Y, Iwao K, Nagasawa Y, Aihara T, Sasaki Y, Imaoka S, Murata M, Shimano T, Nakamura Y (1998) Activation of the beta-catenin gene in primary hepatocellular carcinomas by somatic alterations involving exon 3. *Cancer Res* **58**(12): 2524–2527.
- Morisawa T, Marusawa H, Ueda Y, Iwai A, Okazaki IM, Honjo T, Chiba T (2008) Organ-specific profiles of genetic changes in cancers caused by activation-induced cytidine deaminase expression. *Int J Cancer* **123**(12): 2735–2740.
- Mussil B, Suspene R, Aynaud MM, Gauthier A, Vartanian JP, Wain-Hobson S (2013) Human APOBEC3A isoforms translocate to the nucleus and induce DNA double strand breaks leading to cell stress and death. *PLoS One* **8**(8): e73641.
- Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, Jones D, Hinton J, Marshall J, Stebbings LA, Menzies A, Martin S, Leung K, Chen L, Leroy C, Ramakrishna M, Rance R, Lau KW, Mudie LJ, Varela I, McBride DJ, Bignell GR, Cooke SL, Shlien A, Gammle J, Whitmore I, Maddison M, Tarpey PS, Davies HR, Papaemmanuil E, Stephens PJ, McLaren S, Butler AP, Teague JW, Jonsson G, Garber JE, Silver D, Miron P, Fatima A, Boyault S, Langerod A, Tutt A, Martens JW, Aparicio SA, Borg A, Salomon AV, Thomas G, Borresen-Dale AL, Richardson AL, Neuberger MS, Futreal PA, Campbell PJ, Stratton MR (2012) Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**(5): 979–993.
- Pleasant ED, Cheatham RK, Stephens PJ, McBride DJ, Humphray SJ, Greenman CD, Varela I, Lin ML, Ordóñez GR, Bignell GR, Ye K, Alipaz J, Bauer MJ, Beare D, Butler A, Carter RJ, Chen L, Cox AJ, Edkins S, Kokko-Gonzales PI, Gormley NA, Grocock RJ, Haudenschild CD, Hims MM, James T, Jia M, Kingsbury Z, Leroy C, Marshall J, Menzies A, Mudie LJ, Ning Z, Royce T, Schulz-Trieglaff OB, Spiridou A, Stebbings LA, Szajkowski L, Teague J, Williamson D, Chin L, Ross MT, Campbell PJ, Bentley DR, Futreal PA, Stratton MR (2010) A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**(7278): 191–196.
- Refsland EW, Stenglein MD, Shindo K, Albin JS, Brown WL, Harris RS (2010) Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction. *Nucleic Acids Res* **38**(13): 4274–4284.
- Sheehy AM, Gaddis NC, Choi JD, Malim MH (2002) Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418**(6898): 646–650.
- Shinohara M, Ito K, Shindo K, Matsui M, Sakamoto T, Tada K, Kobayashi M, Kadowaki N, Takaori-Kondo A (2012) APOBEC3B can impair genomic stability by inducing base substitutions in genomic DNA in human cells. *Sci Rep* **2**: 806.
- Stenglein MD, Burns MB, Li M, Lengyel J, Harris RS (2010) APOBEC3 proteins mediate the clearance of foreign DNA from human cells. *Nat Struct Mol Biol* **17**(2): 222–229.
- Stephens PJ, McBride DJ, Lin ML, Varela I, Pleasant ED, Simpson JT, Stebbings LA, Leroy C, Edkins S, Mudie LJ, Greenman CD, Jia M, Latimer C, Teague JW, Lau KW, Burton J, Quail MA, Swerdlow H, Churcher C, Natrajan R, Sieuwerts AM, Martens JW, Silver DP, Langerod A, Russnes HE, Foekens JA, Reis-Filho JS, van't Veer L, Richardson AL, Borresen-Dale AL, Campbell PJ, Futreal PA, Stratton MR (2009) Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* **462**(7276): 1005–1010.

- Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, Nik-Zainal S, Martin S, Varela I, Bignell GR, Yates LR, Papaemmanuil E, Beare D, Butler A, Cheverton A, Gamble J, Hinton J, Jia M, Jayakumar A, Jones D, Latimer C, Lau KW, McLaren S, McBride DJ, Menzies A, Mudie L, Raine K, Rad R, Chapman MS, Teague J, Easton D, Langerod A, Lee MT, Shen CY, Tee BT, Huimin BW, Brooks A, Vargas AC, Turashvili G, Martens J, Fatima A, Miron P, Chin SF, Thomas G, Boyault S, Mariani O, Lakhani SR, van de Vijver M, van 't Veer L, Foekens J, Desmedt C, Sotiriou C, Tutt A, Caldas C, Reis-Filho JS, Aparicio SA, Salomon AV, Borresen-Dale AL, Richardson AL, Campbell PJ, Futreal PA, Stratton MR (2012) The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486**(7403): 400–404.
- Suspène R, Aynaud M, Guétard D, Henry M, Eckhoff G, Marchio A, Pineau P, Dejean A, Vartanian JP, Wain-Hobson S (2011a) Somatic hypermutation of human mitochondrial and nuclear DNA by APOBEC3 cytidine deaminases, a pathway for DNA catabolism. *Proc Natl Acad Sci USA* **108**(12): 4858–4863.
- Suspène R, Aynaud MM, Koch S, Pasdeloup D, Labetoulle M, Gaertner B, Vartanian JP, Meyerhans A, Wain-Hobson S (2011b) Genetic editing of herpes simplex virus 1 and Epstein–Barr herpesvirus genomes by human APOBEC3 cytidine deaminases in culture and *in vivo*. *J Virol* **85**(15): 7594–7602.
- Suspène R, Aynaud MM, Vartanian JP, Wain-Hobson S (2013) Efficient deamination of 5-methylcytidine and 5-substituted deoxycytidine residues in DNA by human APOBEC3A cytidine deaminase. *PLoS One* **8**(8): e73641.
- Suspène R, Guétard D, Henry M, Sommer P, Wain-Hobson S, Vartanian JP (2005b) Extensive editing of both hepatitis B virus DNA strands by APOBEC3 cytidine deaminases *in vitro* and *in vivo*. *Proc Natl Acad Sci USA* **102**(23): 8321–8326.
- Suspène R, Henry M, Guillot S, Wain-Hobson S, Vartanian JP (2005a) Recovery of APOBEC3-edited human immunodeficiency virus G → A hypermutants by differential DNA denaturation PCR. *J Gen Virol* **86**(Part 1): 125–129.
- Suspène R, Sommer P, Henry M, Ferris S, Guétard D, Pochet S, Chester A, Navaratnam N, Wain-Hobson S, Vartanian JP (2004) APOBEC3G is a single-stranded DNA cytidine deaminase and functions independently of HIV reverse transcriptase. *Nucleic Acids Res* **32**(8): 2421–2429.
- Vartanian JP, Guétard D, Henry M, Wain-Hobson S (2008) Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. *Science* **320**(5873): 230–233.
- Vartanian JP, Henry M, Marchio A, Suspène R, Aynaud MM, Guétard D, Cervantes-Gonzalez M, Battiston C, Mazzaferro V, Pineau P, Dejean A, Wain-Hobson S (2010) Massive APOBEC3 editing of hepatitis B viral DNA in cirrhosis. *PLoS Pathogen* **6**(5): e1000928.
- Wang Z, Mosbaugh DW (1988) Uracil-DNA glycosylase inhibitor of bacteriophage PBS2: cloning and effects of expression of the inhibitor gene in *Escherichia coli*. *J Bacteriol* **170**(3): 1082–1091.
- Wijesinghe P, Bhagwat AS (2012) Efficient deamination of 5-methylcytosines in DNA by human APOBEC3A, but not by AID or APOBEC3G. *Nucleic Acids Res* **40**(18): 9206–9217.
- Zhang T, Cai J, Chang J, Yu D, Wu C, Yan T, Zhai K, Bi X, Zhao H, Xu J, Tan W, Qu C, Lin D (2012) Evidence of associations of APOBEC3B gene deletion with susceptibility to persistent HBV infection and hepatocellular carcinoma. *Hum Mol Genet* **22**(6): 1262–1269.

This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported License.

Supplementary Information accompanies this paper on British Journal of Cancer website (<http://www.nature.com/bjc>)

1 SOUMIS

2

3

4 Self cytoplasmic DNA up-regulates the mutator enzyme APOBEC3A
5 leading to chromosomal DNA damage

6

7

8

9 Bianka Mussil^{1,3,#}, Rodolphe Suspène^{1,#}, Hélène Laude¹, Vincent Caval¹,
10 Mohamed S Bouzidi¹, Valérie Thiers¹, Michel Henry¹, Anne Durandy²,
11 Simon Wain-Hobson¹ and Jean-Pierre Vartanian^{1*}

12

13

14 ¹Molecular Retrovirology Unit, Institut Pasteur, 28 rue du Dr. Roux, F-75724 Paris cedex 15

15

16 ²INSERM UMR 1163, The Human Lymphohematopoiesis Laboratory, Institut Imagine,
17 24 boulevard du Montparnasse 75015 Paris

18

19 ³Present address: Unit of Infection Models, German Primate Centre, Goettingen, Germany

20

21 [#] These authors contributed equally to this work

22

23 * Correspondence should be addressed to Jean-Pierre Vartanian

24 Molecular Retrovirology Unit, Institut Pasteur, 28 rue du Dr. Roux, F-75724 Paris cedex 15

25 email, jean-pierre.vartanian@pasteur.fr

26 Phone +33 (0)1 44 38 94 45, Fax +33 (0)1 45 68 88 74

27

28

29 Short title: APOBEC3 catabolism of cytoplasmic DNA agonist

30

31

32

33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51

Abstract

Foreign and self cytoplasmic DNA is recognized by a number of DNA sensor molecules leading to type I interferon production. The DNA agonist should be degraded otherwise cells would be chronically stressed. Most human APOBEC3 cytidine deaminases can initiate catabolism of cytoplasmic mitochondrial DNA (cymtDNA). It is shown here that up to 17% of primary human CD4 T cells show signs of massive APOBEC3 catabolism of cymtDNA. Using the human myeloid cell line THP-1 with an interferon inducible *APOBEC3A* gene we show that cytoplasmic DNA triggers interferon production through the RNA polymerase III transcription/RIG-1 pathway leading to massive up-regulation of *APOBEC3A*. The enzyme initiates catabolism of single stranded DNA preventing such fragments from re-annealing so attenuating the danger signal. The price to pay is chromosomal DNA damage in the form of CG->TA mutations and double stranded DNA breaks which, in the context of chronic inflammation, could drive cells down the path towards cancer.

52 **Introduction**

53 The human *APOBEC3* (*A3*) locus on chromosome 22 is made up of seven genes (*A3A*, *A3B*,
54 *A3C*, *A3DE*, *A3F*, *A3G* and *A3H*) encoding cytidine deaminases with single stranded DNA
55 (ssDNA) specificity [1]. This leaves DNA peppered by uracil residues, the product of cytidine
56 deamination, the process being referred to as genetic editing as it occurs post replication. Uracil
57 bases in DNA are excised by uracil N-glycosidase (UNG) that initiates DNA damage
58 responses, which can result in either DNA repair or catabolism. The *A3* locus arose by
59 duplication and subsequent expansion of the activation-induced deaminase (*AID*) gene in
60 placental mammals [2]. The antiviral role of these A3 enzymes was initially highlighted by
61 their impact on HIV and HBV replication [3-6] and the fact that several were up-regulated by
62 interferon- α [7-9]. A3 enzymes leave a signature in edited DNA: most A3 enzymes
63 preferentially edit a C residue in the context of TpC with the exception of A3G, which prefers
64 CpC dinucleotides. By contrast AID prefers GpC [10].

65
66 A3 enzymes were shown to restrict the transposition of retroelements like SINEs and LINEs
67 [11, 12] while human mitochondrial DNA (mtDNA) in cell lines and tissue shows the
68 hallmarks of A3 editing – TpC and CpC are targetted. The A3 edited mtDNA was cytoplasmic,
69 in keeping with the observation that no A3 enzyme can target the mitochondrial network [13].
70 More recently, A3A and to a lesser extent A3B, were shown to non-specifically target
71 chromosomal DNA leading to hypermutated DNA with up to 70% of cytidine residues
72 deaminated [13, 14]. Not surprisingly such levels of A3A-induced DNA damage led to the
73 formation of double stranded DNA breaks (DSB), cell cycle arrest and apoptosis [15, 16]. The
74 cellular location of A3A is both cytoplasmic and nuclear while A3B is exclusively nuclear [14,
75 15]. Despite this, A3A is probably the more active of the two enzymes [14]. Their role in
76 cancer is highlighted by the fact that many cancer genomes encode tens of thousands of C->T
77 transitions frequently in the TpC dinucleotide context, a hallmark of A3A and A3B
78 deamination [17-19]. Epidemiological studies have shown that an *A3B* deletion polymorphism,
79 whereby all but the last exon of *A3B* is deleted and correlated with a higher odds ratio of
80 developing breast, ovarian or HBV-associated liver cancer [20-23]. Breast cancer genomes
81 from $\Delta A3B^{-/-}$ patients harbor a higher mutation burden [24] while the chimeric *A3A* transcript
82 resulting from the *A3B* deletion is more stable leading to greater intracellular steady state levels
83 of A3A [14]. In so doing, these studies indicate that A3A alone can generate human cancers,
84 and show that the $\Delta A3B$ lesion is a causal cancer susceptibility marker [14, 24].

Cancer emerges on a background of chronic inflammation [25], those associated with chronic hepatitis B, hepatitis C virus (HBV and HCV) and *Helicobacter pylori* infections being well-known. Toll like receptors (TLRs) recognize microbial components, for example CpG DNA and TLR9, resulting in IFN α production in plasmacytoid dendritic cells, [26]. As mitochondrial DNA fragments (mtDNA) resembles bacterial DNA in possessing unmethylated CpG motifs, it too can trigger cytoplasmic DNA sensor molecules leading to inflammatory responses [27]. Recent work showed that mtDNA can induce TLR9-mediated inflammatory responses in cardiomyocytes and is even capable of inducing myocarditis [28]. Cellular disruption can release mtDNA into the circulation causing systemic inflammation [29], while extracellular DNA can be taken up into endosomes and sensed by specific TLRs [26].

Intracellular DNA can be captured by a myriad of DNA sensor molecules leading to the triggering of potent innate immune responses [30]. The DNA cytosolic binding protein LRRFIP1 interacts with β -catenin and promotes its activation by phosphorylation [31]. After binding to the C-terminal domain of the transcription factor IRF-3, IFN β production is initiated [31]. cGAS can detect cytoplasmic DNA and induce IFN [32]. Equally, AT-rich DNA can trigger type I IFN responses via intracytoplasmic transcription of double stranded DNA (dsDNA) by RNA polymerase III to form a dsRNA intermediate sensed by RIG-I [33, 34]. Interestingly, these different pathways have in common STING (Stimulator of Interferon gene) that was shown to be pivotal for the production of type I IFNs [35]. STING itself interacts with both MAVS and RIG-I, themselves crucial to interferon signaling [36].

Despite the wealth of knowledge concerning cytoplasmic DNA signaling, there is much less data on the fate of the DNA agonist, for the danger signal has to be countered. This is suggested by the fact that germ line mutations in some endo- and exonuclease genes result in symptoms resembling chronic inflammation [37, 38]. We have previously shown that cytoplasmic mtDNA (cymtDNA) can be deaminated by the human A3 enzymes [13]. While most effete or spent mitochondria are phagocytosed, some mtDNA fragments clearly find their way to the cytoplasm so acting as a danger signal. Patients with lesions in the *UNG* gene showed much higher levels of deaminated cymtDNA indicating that the A3-targeted mtDNA was earmarked for catabolism [13].

In the present study, we analyzed the ability of A3 deamination of cymtDNA at a single cell level. Up to 25% of cells involving hundreds of edited mtDNA genomes were shown to be

hyperedited. In the human monocytic cell line THP-1, transfected dsDNA was transcribed by RNA polymerase III leading to dsRNA intermediates that were captured by RIG-I ultimately leading to type I interferon production. In turn A3 cytidine deaminases were induced, particularly A3A which resulted in deamination and catabolism of the transfected DNA agonist. The data highlight the A3 cytidine deaminases as anti-inflammatory agents. The sting in the tail is A3A mediated chromosomal DNA damage.

Results

Editing of cymtDNA at the single cell level

CymtDNA remains the most sensitive target to detect A3 editing. We first analyzed P2 cells, an EBV transformed B-cell line from an UNG deficient individual, which reveals more readily A3 activity [13]. Viable single P2 cells were FACS sorted into 96-well PCR plates. Over 500 cells were analyzed by an established PCR/3DPCR procedure [39] specific for part of the *MT-COI* gene (mitochondrial cytochrome c oxidase subunit I) to identify cells harboring A3-edited mtDNA [13]. Among 512 P2 cells only three (0.6%) proved positive (Fig 1A). First round *MT-COI* PCR products from cells B02 and E11 as well as those from a negative cell (D05) were used for 3DPCR using a gradient of denaturation temperatures. As can be seen from Fig 2A, *MT-COI* DNA could be recovered down to 81.6°C and 84.6°C for cells B02 and E11. By contrast no mtDNA was recovered below 86.7°C with D05 cell, the lowest temperature at which normal *MT-COI* DNA can be amplified. Cloning and sequencing of 3DPCR products identified by an asterisk (Fig 2A) showed massive editing of *MT-COI* minus strand DNA, ~40-50 fold over background (cell D05, Fig 2B). The dinucleotide context associated with cytidine editing showed a strong preference for TpC (or GpA on the other strand) and CpC (or GpG) (Figs 2C and 2D) consistent with A3 deamination. Single molecule analysis showed that most were edited at TpC and CpC sites even though locally some regions were exclusively edited at TpC or CpC sites, consistent with a degree of processivity typical of these enzymes (Fig 2E).

To get an idea as to which human A3 enzymes were involved, quail QT6 cells were transfected by A3 expression plasmids and the equivalent region of quail mtDNA analyzed. As the quail genome does not encode any A3 gene, there is no endogenous background editing. A3A, A3C, A3F and A3G could all access quail mtDNA as previously reported [13]. Comparisons of the dinucleotide context with those from bulk P2 cells (Fig 2F) as well as single molecule analyses indicate that A3A and A3C are the more plausible candidates (Fig 2G). Human A3B does not

153 edit quail mtDNA presumably because it is too rapidly translocated to the nucleus.

154

155 **Editing of cymtDNA in primary CD4⁺ T lymphocytes**

156 As numerous studies have described substantial *A3* expression in leukocytes [9, 40], we
157 addressed a more physiological situation. CD4⁺ T lymphocytes were purified from two healthy
158 donors (D1 & D2). Ficoll purified non-stimulated CD4⁺ T lymphocytes were collected after
159 overnight culture and FACS sorted. Viable single cells were analyzed as per P2 cells (Fig 1A).
160 When analyzed by PCR at Td=85°C ~17% and 12% of cells from D1 and D2 harbored
161 hyperedited *MT-COI* DNA (Fig 1A), ~20-28 fold greater than for P2 cells. Two single cells
162 from D1 (A06 and F09, S1A Fig) and D2 (C08 and E02, S2A Fig) were analyzed in detail.
163 3DPCR recovered edited *MT-COI* DNA far below the reference temperature of ~86.7°C (S1A
164 and S2A Figs). Mutation matrices (S1B and S2B Figs) and sequence analyses highlighted
165 hyperediting by A3 enzymes, again with a strong bias in favor of TpC and CpC as typical for
166 A3 deaminases (S1C and S2C Figs). To assess the relative expression levels of *A3* genes, a
167 TaqMan transcriptional study was made of the seven *A3* genes using *RPL13A* as reference. The
168 primary CD4⁺ T cells showed significantly higher *A3A* and lower *A3B* levels compared to P2
169 cells suggesting that A3A activity might account for the greater proportion of primary cells
170 harboring edited mtDNA (S1D and S2D Figs).

171

172 **Etoposide and actinomycin D induced apoptosis and *A3* expression**

173 Given the low frequency of P2 cells harboring hyperedited cymtDNA, perhaps linked to the
174 relatively low levels of A3A transcripts, we tested the impact of genotoxic stress on cymtDNA
175 deamination. Etoposide is a topoisomerase II inhibitor while actinomycin D binds to DNA and
176 inhibits transcription initiation resulting in DSBs leading to apoptosis [16]. Treatment by these
177 molecules resulted in cytochrome c release and apoptosis at 16 hours (Figs 3B and 3C). An *A3*
178 transcriptome analysis showed up-regulation of *A3A*, *A3F*, *A3G* and *A3H* for etoposide and *A3A*
179 after actinomycin D treatment using a robust >3 fold increase as cut off (Fig 3A).

180

181 After treatment with 100 µM etoposide or actinomycin D for 16 hours viable P2 cells were
182 individually FACS sorted into 96-well PCR plates and the proportion of cells harboring
183 hyperedited mtDNA at Td=85°C was determined as above. Treatment considerably increased
184 the proportion ~40 fold (Fig 1A). First round PCR products from two cells treated with 100 µM
185 etoposide (D08 and E10) or 100 µM actinomycin D (A04 and F06) were analyzed in detail as
186 described above (Fig 3D) and compared to untreated C08 cell. PCR products obtained at

86.1°C were cloned and sequenced. Minus strand mtDNA was heavily hyperedited (Fig 3F), the dinucleotide context being in favor of TpC and CpC (Fig 3E) indicating the involvement of A3 enzymes in the genotoxic stress phenotype associated with etoposide and actinomycin D.

As 3DPCR preferentially recovers heavily as opposed to lightly deaminated DNA it is not a very quantitative technique [39]. To identify lightly deaminated cytmDNA, first round PCR DNA recovered at 95°C from two cells from each sample identified in Fig 1A were cloned and sequenced. While this data concerns mtDNA located both in the mitochondria and cytoplasm interestingly ~50% harbored between 1-5 CG->AT transitions while 1-4% were hypermutated in most samples (Fig 1B) and showed a TpC context. Hypermutated sequences obtained at 95°C were heavily mutated with both strands being vulnerable (Fig 1C). The hypomutated sequences show that the proportions of cells harboring A3 hyperedited cytmDNA shown in Fig 1A are clearly underestimates. Possibly most amplifiable cytmDNA fragments suffer from cytidine deamination.

Sepsis and circulating edited mtDNA

In view of an experimental link between cell stress and cytoplasmic mtDNA editing (Fig 3), we sought to establish this in a more physiological setting. Sepsis is generally characterized by systemic inflammation due to microbial invasion of the bloodstream [41]. The same nested PCR/3DPCR approach was used to recover *MT-COI* DNA from DNA extracted from the serum from 11 patients with sepsis [42] compared to ten healthy donors (Fig 4A). *MT-COI* DNA was recovered at denaturing temperatures as low as 83.1°C from 10/11 (~90%) samples tested, well below that for *MT-COI* DNA (Td-86.7°C, Fig 4A). By contrast 4/10 healthy serum controls showed signs of editing (Fig 4A). Molecular cloning and sequencing confirmed that they represented A3 edited *MT-COI* DNA with the classic TpC and CpC editing bias (Fig 4B). No correlation was found between the degree of editing and serum IL6 levels. Since the proportion of positive samples from patients with sepsis was greater than controls findings, it could reflect the highly inflammatory microenvironment and release of necrotic cells into the periphery.

Uracil in cytosolic DNA is not a danger signal

Given that hyperedited cytmDNA bears non-canonical uridine residues, they might constitute a novel danger signal. To explore this hypothesis, *MT-COI* DNA was amplified using either dUTP, an equimolar mix of dTTP and dUTP, or just dTTP. To control for the incorporation of dUTP in the DNA, a second round of PCR was performed with internal primers using *Pfu* or *Taq* DNA polymerase (Fig 5A). Like all archaeal DNA polymerases, *Pfu* is unable to amplify

221 DNA templates bearing dU. As expected *Pfu* PCR failed to recover DNA when dUTP (*MT-*
222 *COI_U*) or mixed nucleotides dUTP+dTTP (*MT-COI_{UT}*) were incorporated, in contrast to *Taq*
223 PCR (Fig 5A).

224
225 THP-1 is an IFN α -inducible human acute monocytic leukemia cell line that is one of the very
226 few where all *A3* genes are expressed or can be induced, particularly *A3A*. THP-1 cells were
227 transfected by *MT-COI_T*, *MT-COI_U* or *MT-COI_{UT}* DNA. Supernatants were clarified at 24 h and
228 IFN α and - β production were measured by ELISA. As can be seen in Figs 5B and 5C, 500 ng
229 of *MT-COI_T*, *MT-COI_U* or *MT-COI_{UT}* generated equally strong IFN responses compared to the
230 jetPRIME control indicating that uridine in DNA did not constitute a novel danger signal.
231 Moreover, THP-1 cells incubated with 500 ng of *MT-COI_T* DNA (*MT-COI_T**), but not
232 transfected, failed to induce IFN α or IFN β , indicating that intracellular DNA constitutes the
233 trigger (Figs 5B and 5C). As THP-1 cells are insensitive to unmethylated CpG and polyIC, IFN
234 induction is not via TLR3 and TLR9 pathways [43] (Figs 5B and 5C).

235
236 Treatment of THP-1 cells by *MT-COI_T* DNA increased *A3A*, *A3F*, *A3G* and *A3H* expression at
237 24 hours in a dose dependent manner, even at concentrations that did not induce detectable
238 IFN α or IFN β levels in culture supernatants (Fig 5D). *A3A* was by far the most sensitive to
239 transfected DNA being upregulated by several logs (Fig 5D). As similar *A3* transcription
240 profiles were observed using *MT-COI_U* or *MT-COI_{UT}* DNA (S3A and S3B Figs) this again
241 suggests that uracil in DNA does not constitute a novel danger signal leading to IFN α induction
242 and up-regulation of *A3A*.

244 ***MT-COI* DNA activates interferon leading to *A3* expression**

245 In order to see if exogenous DNA induction of *A3* genes could edit endogenous cytmDNA,
246 THP-1 cells were transfected with *MT-COI_T* DNA and cytochrome b mtDNA (*MT-CYB*) was
247 amplified as previously described [13]. The 3DPCR protocol recovered DNA down to lower
248 Tds than the jetPRIME control, indicative of increased *A3* editing (Fig 5E). 3DPCR products
249 recovered at 80.4°C were cloned and sequenced. The sequences were peppered by C->T
250 substitutions, the mean editing frequency per sequence increasing with the quantity of
251 transfected DNA (Fig 5F).

252
253 Given the very low transfection frequency of THP-1 cells by plasmid DNA, <1% using a GFP
254 expression plasmid, up regulation of *A3A* by several logs presumably was mediated by a small

number of transfected cells and amplified by paracrine IFN α / β induction via IFNAR. With this in mind, non-stimulated PBMCs, which are notoriously difficult to transfect efficiently, were transfected with *MT-COI_T* DNA. IFN α induction was comparable to that for THP-1 (Fig 5B vs. S4A), although IFN β production was ~40 fold lower (Fig 5C vs. S4B). Again, DNA had to be transfected in order to obtain IFN induction (*MT-COI_{T, U, UT}*). The strength of IFN induction again indicates amplification presumably via the paracrine effect. Indeed, treatment of THP-1 cells by IFN α for 16h or 24h increased A3A expression in a dose dependent manner (S4C Fig).

In order to distinguish whether ssDNA or dsDNA was the IFN agonist, four pairs of complementary oligonucleotides corresponding to a small region of *MT-COI* were synthesized - WT₁₊₂ (43 bp, 51% GC) and WT₃₊₄ (39 bp, 56% GC) while oligos Hyp₁₊₂ (43 bp, 7% GC) and Hyp₃₊₄ (39 bp, 13% GC) correspond to hypermutated forms of WT₁₊₂ and WT₃₊₄ respectively (Table S1). THP-1 cells were transfected with heteroduplexes and IFN α induction measured (Fig 6A). WT₁₊₂ and WT₃₊₄ (dsDNA) were potent inducers of IFN α , and on a par with the 511 bp *MT-COI* DNA fragment, whereas the mismatched WT₁₊₃ and WT₂₊₄, effectively ssDNA, failed to induce IFN α (Fig 6A). As the complementary oligonucleotides Hyp₁₊₂ and Hyp₃₊₄ did not induce IFN expression, it is possible that they did not form stable heteroduplexes at 37°C. These findings strongly suggest that dsDNA was necessary for IFN induction. Indeed, *MT-COI* DNA transfection of THP-1 cells led to significantly increased DSBs (Fig 6B) as expected from IFN induced *A3A* up-regulation (Fig 5D).

To focus on the transfected cells alone, THP-1 cells were transfected with the same small DNA molecules labeled with a FAM-fluorophore, WT₁₊₂FAM (dsDNA) or WT₃₊₂FAM (ssDNA). Twenty-four hours post-transfection, addition of FAM to DNA did not induce DSB (WT₃₊₂ vs. WT₃₊₂FAM and WT₁₊₂ vs. WT₁₊₂FAM, Fig 6C). However, as FAM-fluorescence decreased over time, FAM⁺ positive cells were sorted six hours post-transfection and cultured for a further 18 hours and fixed for DSB FACS analysis. Transfection of the double-stranded probe led to ~40% of γ H2AX positive THP-1 cells (Fig 6D) which was correlated with *A3A* up-regulation (Fig 6E).

Cytosolic dsDNA is transcribed by RNA polymerase III and triggers the RIG-I pathway

Which of the cytosolic DNA sensing pathway controls the type I IFN response? At 48 hours post-transfection of THP-1 cells by *MT-COI* DNA, a Western blot analysis showed increased levels of the phosphorylated forms of TBK1 (TBK1-P) and IRF3 (IRF3-P) as well as DNA

289 dependent increases in RIG-1 and MDA5 (Fig 7A). Steady state levels of STING, IKK ϵ ,
290 MAVS were unchanged. Dose dependent increases of TBK1-P and IRF3-P, key regulators of
291 IFN induction, suggest a mechanism by which IFN α and IFN β were induced.

292
293 As RNA polymerase III has been implicated in the transcription of cytosolic dsDNA [33, 34],
294 THP-1 cells were transfected with *MT-COI_T* or *MT-COI_U* DNA along with ML-60218, an
295 inhibitor of RNA polymerase III [44] (Fig 7B). There was an inverse correlation between IFN α
296 production and inhibitor concentration. In order to exclude drug toxicity, 20 μ g/ml of
297 peptidoglycans (PGN) known to stimulate IFN α via TLR2, was used. IFN production was
298 insensitive to all inhibitor concentrations used (Fig 7B). In addition qRT-PCR of *RPL13A* was
299 unaffected by the drug, again demonstrating no toxicity at the concentration used. The inhibitor
300 abolished transfected *MT-COI_T* DNA induced A3 mediated hyperediting of endogenous
301 cytoplasmic *MT-CYB* DNA, again something that treatment with PGN and ML-60218 alone or
302 in combination failed to do (Fig 7C).

303
304 If cymtDNA was signaling via the RIG-I pathway downstream of RNA Polymerase III
305 transcription, it should be possible to identify such transcripts. In the previous experiments a
306 strong background would come from endogenous mtRNA transcripts. To overcome this, THP-
307 1 cells were transfected by PCR products corresponding to a segment of HIV-1 envelope
308 (V1V2, 685 bp, 38% GC [45]). THP-1 cells transfected with 500 ng of V1V2 DNA induced
309 IFN α/β to a degree comparable to *MT-COI_T* (Fig 7D vs. Figs 5B and 5C). Total RNA was
310 extracted from two independent transfections, treated with DNase and RT-PCR performed.
311 V1V2 RNA intermediates were recovered only when a reverse transcriptase step was
312 performed (Fig 7E). Lysates were made 8 hours post transfection and immunoprecipitation
313 with an anti-RIG-I monoclonal antibody was performed along with an anti-HA as control.
314 V1V2 specific RT-PCR products were recovered only from the anti-RIG-I immunoprecipitates
315 (Fig 7F).

316 Discussion

317 The APOBEC3 cytidine deaminases are particularly well expressed in hematopoietic cells [9,
318 46, 47]. They can edit cymtDNA, which is coherent with a cytoplasmic or nucleo-cytoplasmic
319 localization; only A3B failed to do so, presumably because it is rapidly translocated to the
320 nucleus. The fraction of cells harboring A3-hyperedited cymtDNA varied from 0.6-25% and
321 could be increased by genotoxic stress (Fig 1A) while the number of hyperedited cymtDNA
322 fragments per cell is of the order of 1-4% (Fig 1B). The mitochondrial network varies

considerably between resting and proliferating cells, the number of mtDNA genomes varying, ranging from 3-10K mtDNA genomes per cell. Accordingly the number of A3-hyperedited cymtDNA molecules recovered can be as high as $(3-10K) \times (1-4\%) = 30-400$ per cell at any moment. Interestingly the proportion of *MT-COI* sequences harboring 1-5 CG->TA mutations was consistently ~50%. As the pool of mtDNA sequences was derived from the cytoplasm and mitochondrial network, this suggests that in these cells approximately half of mtDNA genomes, thousands of copies, are fragmented in the cytoplasm for A3 enzymes do not access the mitochondrion [13] and can only deaminate ssDNA. As up to 17% of unstimulated primary CD4⁺ T cells from blood donors harbor hypermutated cymtDNA, what does this tell us? The fraction is too large to reflect proliferating effector cells in the periphery so they may reflect recently activated memory cells. Even so, the fraction is too large to reflect an antigen or pathogen specific population of cells.

Transfected PCR products can be captured by RNA polymerase III leading to transcription and annealing of RNA products to form dsRNA followed by RIG-I capture and signaling (Fig 7). Single stranded DNA signaling was weak by comparison (Fig 6A). Presumably, DNA heteroduplexes as small as 39 bp can also be transcribed for IFN α/β induction is almost as efficient as for the 511 bp *MT-COI* fragment (Fig 6A). These findings do not exclude other pathways in primary cells for THP-1 cells do not mount TLR3 and TLR9 dependent responses. However, as the ML-60218 inhibitor completely blocks IFN α and β induction, RNA polymerase III is a major danger signaling pathway in these cells. For the THP-1 cells (from the ATCC) used in this study the *A3A* gene could be induced to very high levels, just like PBMCs suggesting that it could be an important pathway in primary cells.

Although A3 enzymes catalyze oxidation of cytidine to uridine in ssDNA, the uracil base does not itself constitute a novel danger signal per se (Figs 5B and 5C). However, as AU-rich edited DNA is thermodynamically less stable it represents a means to prevent annealing of complimentary ssDNA fragments arising from exonuclease activity, which itself reduces the agonist effect of dsDNA (Fig 6A). This is particularly relevant for cymtDNA. If hundreds of mtDNA genomes can leak out to the cytoplasm and be processed by dsDNA endonucleases followed by dsDNA exonucleases such as TREX1 which leaves ssDNA there is a distinct probability that complementary single stranded mtDNA molecules could re-anneal to dsDNA and re-signal danger via DNA sensor molecules.

357 Cytidine deamination of ssDNA is rapidly followed by removal of uracil moieties by the highly
358 efficient enzyme UNG followed by cleavage of the ssDNA by apurinic/apyrimidinic
359 endonucleases such as APE1 and 2. Together the three are equivalent to a cytidine specific
360 endonuclease. The fact that A3-edited mtDNA is found in so many situations - fresh PBMCs,
361 THP-1 cells, etoposide and actinomycin D treated P2 cells, circulating DNA in healthy patients
362 as well as those with sepsis, while an orthologous A3A enzyme is conserved across 150 MYr
363 of evolution [48] - suggests that the cytoplasm regularly harbors ssDNA fragments requiring
364 catabolism. In this context, it is interesting that Aicardi-Goutières patients with a *TREX1* lesion
365 genes show signs of chronic inflammation [37, 49]. Presumably, APOBEC3 cytidine
366 deaminases and TREX1 act in concert to catabolize cytoplasmic DNA whatever its origin. The
367 higher frequency of edited *MT-COI* DNA present in the serum of sepsis patients (10/11, Fig
368 4A) compared to healthy patients (4/10, Fig 4A) suggests that their proportion increases with
369 cell stress and/or inflammation, although there was no correlation with serum IL6 levels (Fig
370 4B).

371
372 The findings highlights a mechanism by type I IFN induced A3A expression leads to
373 catabolism of the DNA agonist. Disruption of parts of the mitochondrial network, perhaps as a
374 result of endoplasmic reticulum, oxidative or genotoxic stress, results in release of mtDNA to
375 the cytosol. Double-stranded DNA serves as a template for cytoplasmic RNA synthesis by
376 DNA-dependent RNA polymerase III. The complementary strands which anneal to yield
377 dsRNA recognized by RIG-I pathway (Fig 8). This leads to phosphorylation of IRF3 by TBK1
378 which, upon dimerization and translocation into the nucleus, will activate transcription of type I
379 interferon. Interferon in turns activates several *A3* genes leading to heightened catabolism of
380 single stranded DNA generated by cytoplasmic exonucleases (Fig 8).

381
382 There could be two steps in the catabolism of cytoplasmic ssDNA. A3C, A3F and A3G, which
383 can edit cymtDNA, are expressed to some degree in most cells [1, 13] and presumably may
384 function as constitutive catabolic enzymes. By contrast, A3A (nucleo-cytoplasmic) is massively
385 upregulated by IFN α as is A3G (cytoplasmic) to a lesser extent [9]. When THP-1 cells are
386 triggered by high levels of cytoplasmic DNA, not only cymtDNA becomes more edited (Fig
387 5F) but also the DNA shows signs of more editing in the dinucleotide context TpC typical of
388 A3A rather than CpC, characteristic of A3G, suggesting that A3A outclasses A3G.

389

The findings tie in well with the massive egress of mtDNA to the cytosol following a mitochondrial genetic lesion as well as infection stress following herpesvirus infection [32]. Here, the signaling pathway was the DNA sensor cGAS that promoted STING-IRF3 dependent signaling resulting in IFN production. The difference here is that that finding pertains to the mouse, which unusually for mammals, does not encode an A3A ortholog [48]. Although different DNA sensors probably overlap and converge on induction of IFN and a vast array of downstream effector molecules [50] catabolism of the DNA agonist in the mouse may proceed by a different mechanism for we were unable to detect cytidine deaminated mtDNA in primary tissues [13].

There is a price to pay. IFN α induction of *A3A* results in genetic hyperediting of chromosomal DNA (Fig 6, [16]). The link between A3A and cancer has been established epidemiologically [17, 51-53]. Cancer genomes are characterized by tens of thousands to more than a hundred thousand CG->TA transitions, frequently in the tell tale TpC signature, deamination of 5-methylcytidine residues and large numbers of DNA rearrangements [17, 19, 51, 54-56]. The A3A DNA mutator enzyme can reproduce these three mutational hallmarks experimentally [16, 45]. In addition some cancer genomes harbor discrete regions of hypermutation, a process referred to as kataegis, which has been attributed to A3A and A3B activities [51, 57]. A3B can edit C and 5MeC in single stranded nuclear DNA although DSB formation, which presumably occurs, is below the detection threshold. From an experimental setting, A3A appears to be the more active of the two enzymes [14]. The major difference is that A3A is massively upregulated by IFN- α in hematopoietic cells while A3B is not [9].

Chronic inflammation, autoimmune diseases such as systemic lupus erythematosus and interferonopathies are associated with an increased risk for cancer. All show type I interferon signatures of which A3A is part. While a powerful catabolic enzyme in the cytoplasm A3A can locate to the nucleus and turn on nuclear DNA so setting up mutation/selection among somatically edited cells. Through the interferon paracrine effect, bystander cells too could undergo somatic mutation. It would be interesting to explore A3A-mediated genotoxic damage in other inflammatory pathologies such as diabetes, atherosclerosis, and perhaps even ageing. In conclusion, these findings show that the danger signal triggered by exogenous or endogenous cytosolic DNA encodes a mechanism for catabolism of the DNA agonist. A3A can be seen as anti-inflammatory enzyme.

424

425 **Materials and Methods**

426 **Reagents**

427 Etoposide and dUTP were from Sigma, actinomycin D (Millipore), RNA polymerase III
428 inhibitor (ML-60218) was from Merck Millipore, dNTP (Fermentas), Type I IFN α were from
429 PBL Biomedical Laboratories, CpG (Invitrogen), polyIC and Peptidoglycan (InvivoGen), Pfu
430 (Agilent Technologies), STING (antibody #3337), phospho-IRF-3 (Ser396) (4D4G, rabbit mAb
431 #4947), IRF-3 (D6I4C, XP rabbit mAb #11904), β -catenin antibody (amino-terminal antigen,
432 #9581), mouse anti-rabbit IgG (conformation specific, L27A9, mAb #3678), anti-rabbit IgG,
433 HRP-linked antibody #7074 were from Cell Signaling Technology. RIG-I antibodies (E-5, sc-
434 376882) were from Santa Cruz Biotechnology, Inc. Anti-mouse IgG, (HRP-linked antibody
435 #NA931V) was from GE Healthcare. Monoclonal anti- β -actin-peroxidase (antibody #A3854)
436 was from Sigma Aldrich.

437

438 **Plasmids**

439 The V5-tagged A3A, A3B, A3C, A3DE, A3F, A3G and A3H plasmids in the pcDNA3.1D/V5-
440 His-TOPO expression vector (Invitrogen) have been described [13]. All constructs were
441 transformed and amplified in *E. coli* DH5 α .

442

443 **Cell culture and transfection**

444 The Japanese quail muscle fibroblast cell line QT6 (ATCC[®] CRL-1708[™]) was maintained as
445 described [13]. For transfection 7x10⁵ QT6 cells were seeded in six-well tissue culture plates
446 and incubated for 24 h. Transfections were performed using jetPRIME (Polyplus transfection).
447 At 48 h, DNA was extracted using the MasterPure Complete DNA and RNA purification kit
448 (Epicentre Biotechnologies).

449

450

451 THP-1 cells (ATCC[®] TIB-202[™]) were maintained in RPMI (Eurobio), supplemented with 50
452 U/ml penicillin, and 50 μ g/ml streptomycin, and 10% heat-inactivated fetal calf serum. Cells
453 were grown in 75 cm² cell culture flasks at 37°C in a humidified atmosphere containing 5%
454 CO₂. For transfection 1.5x10⁶ THP-1 cells were seeded in twelve-well tissue culture plates and
455 incubated for 24 hours. DNA transfections were performed using jetPRIME (Polyplus
456 transfection). At 24 h post transfection, supernatants were harvested and INF α (VeriKine[™]
457 Human IFN Alpha Multi-subtype ELISA Kit) and IFN β (VeriKine[™] Human IFN Beta Multi-

subtype ELISA Kit) were quantified and analyzed (PBL Assay Science). All DNAs were extracted using the MasterPure Complete DNA and RNA purification kit (Epicentre Biotechnologies).

Human P2 cells, an EBV transformed B-cell line obtained from an UNG-deficient human patient [58] were maintained as described [13]. Apoptosis was induced by incubating P2 cells with 100 μ M actinomycin D or 100 μ M etoposide for 16 h. Cells were collected and used for DNA extraction as well as for single cell sorting.

PBMC and isolation of CD4⁺ T lymphocytes

Blood was obtained from anonymous healthy donors (Authorisation IP: HS2004-3165 and HS2012-24917) and approved by the Comité Consultatif National d'Ethique (CCNE) de la Direction Générale et d'instances éthique et déontologique de l'Institut Pasteur. Federalwide Assurance (FWA) for the Protection of Human Subjects is FWA00003327 (N° IRB : 00006966). The anonymous healthy donors provided their written informed consent to participate in this study. Peripheral blood mononuclear cells (PBMC) were isolated by Ficoll gradient (Eurobio). Isolation of CD4⁺ T lymphocytes was performed by incubation with antibody-coated magnetic beads (Miltenyi Biotec). Purity of CD4⁺ T lymphocytes was above 90% as checked by flow cytometry (FACSCalibur, Becton Dickinson). Two million CD4⁺ T lymphocytes were seeded in 24-well plates, cultured overnight and subsequently applied for single cell sorting.

Single cell sorting

P2 cells were collected 16 h post-treatment with the apoptosis inducing drugs actinomycin and etoposide. In parallel, untreated P2 cells were harvested. CD4⁺ T lymphocytes were collected after overnight culture. All cells were resuspended in 500 μ l PBS and viable single cells as defined by gating were sorted into 96-well PCR plates (ABgene, Thermo Scientific) using the MoFlo (Beckman Coulter). Plates were immediately sealed with PCR adhesive film (Thermo Scientific), stored at -20°C and analyzed by PCR.

Mitochondrial cytochrome c release

At 16 h post-treatment with apoptosis inducing drugs actinomycin and etoposide, P2 cells were investigated for cytochrome c release using the FlowCollect Cytochrome c Kit from Millipore following manufacturer's instructions. Cells were analyzed with FACSCalibur using Cell Quest

492 Pro (BD Biosciences) or FlowJo software (Tree Star, Inc., version 8.7.1). For each sample
493 10,000 cells were counted.

494

495 **FACS analysis for apoptosis**

496 Annexin V possesses high affinity for the phospholipid phosphatidylserine (PS) thus
497 identifying cells undergoing apoptosis. At 16 h post-treatment with apoptosis inducing drugs
498 actinomycin and etoposide, P2 cells were resuspended in binding buffer (BD Pharmingen) and
499 stained with FITC-labeled Annexin V antibody (1 µg/ml) (BD Pharmingen). Discriminating
500 between early apoptotic and late apoptotic or necrotic events was achieved by counterstaining
501 cells with 5 µg/ml propidium iodide (PI) (BD Pharmingen). Cells were analyzed with
502 FACSCalibur using CellQuest Pro or FlowJo software. For each sample 10,000 events were
503 collected.

504

505 **FACS analysis for DNA double stranded breaks**

506 Twenty-four hours post transfection THP-1 cells were washed with PBS, fixed in 2% ice-cold
507 paraformaldehyde (Electron Microscopy Sciences) for 15 min and permeabilized in 90% ice-
508 cold methanol (Sigma) for 30 min. After two washes with PBS, cells were incubated for 1 h
509 with 1:100 diluted Alexa Fluor 488-conjugated mouse monoclonal anti-γH2AX (N1-431)
510 antibody (BD Pharmingen). All incubation steps were performed on ice. Cells were analyzed
511 on MACSQuant Analyzer (Miltenyi Biotec) using the MACsQuantify™ Software (Miltenyi
512 Biotec) or FlowJo software (Tree Star, Inc., version 8.7.1). For each sample 10,000 cells were
513 counted.

514

515 **PCR/3DPCR**

516 Sepsis Patients have been described [42] and the protocol has been approved by the ethical
517 committee of Pitié-Salpêtrière Hospital, Paris (NCT00698919). First round reaction parameters
518 for human or quail *MT-COI* were 95°C for 5 min, followed by 40 cycles (95°C or 85°C for 30
519 s, 60°C for 30 s, and 72°C for 2 min), and finally 10 min at 72°C. Second round standard PCR
520 and 3DPCR were performed using the equivalent of 1 µl of the first round reaction as input.
521 The reaction parameters for standard PCR were 95°C for 5 min, followed by 40 cycles (95°C
522 for 30 s, 55 or 60°C for 30 s, and 72°C for 2 min), and finally 10 min at 72°C. Second round
523 3DPCR was performed using an Eppendorf gradient Mastercycler S. The reaction parameters
524 were 81 to 87°C for 5 min, followed by 40 cycles (81–87°C for 30 s, 60°C for 30 s, and 72°C
525 for 2 min), and finally 10 min at 72°C. For amplification of human *MT-CYB*, the first-round

reaction parameters were 95°C for 5 min, followed by 40 cycles (95°C for 30 s, 60°C for 30 s, and 72°C for 2 min), and finally 10 min at 72°C. For second-round 3DPCR the reaction parameters were 81 to 87°C for 5 min, followed by 40 cycles (81–87°C for 30 s, 60°C for 30 s, and 72°C for 2 min), and finally 10 min at 72°C. All amplifications were carried out using 2.5 U Taq (Bioline) DNA polymerase per reaction. PCR conditions and primers were described before [13, 39]. HIV-1 V1V2 region amplification was already described [45].

Real-time PCR

Total RNA was extracted from P2 cells, THP-1 and CD4⁺ T lymphocytes using RNeasy Plus Mini Kit (Qiagen) according to the manufacturer's instructions. Synthesis of cDNA was performed with 1 µg RNA using the Quantitec Reverse Transcription Kit (Qiagen). Quantitative PCR was performed using cDNA samples and TaqMan Universal PCR Master Mix (Applied Biosystems) for human A3A to A3H. Primers and PCR condition were described before [59]. Data were normalized to the expression levels of the housekeeping reference gene *RPL13A*.

Western blotting

Total protein was recovered 24 hours post transfection. Western blot analysis was carried out according to standard procedures by using a mouse or rabbit monoclonal specific antibodies. After incubation with an anti-mouse IgG horseradish peroxidase-coupled secondary antibody (Amersham), the membrane was subjected to detection by enhanced chemiluminescence (Pierce).

RNA protein co-immunoprecipitation

Eight hours after 500 ng of HIV-1 V1V2 DNA transfection, 1.5x10⁶ THP-1 transfected cells were incubated with 37% formaldehyde (to a final concentration of 1%) for 15 minutes at room temperature under gentle shaking. Cross-linking was stopped by adding 2 M glycine to a final concentration of 0.2 M for 5 minutes at room temperature. Cells were then centrifuged at 1800 g for 5 minutes at 4°C and lysed in Cell Lysis Buffer (#9803, Cell Signaling) containing 40 U/mL RNasin on ice for 5 minutes. Lysates were then sonicated on ice three times for 5 seconds each. After centrifugation at 14,000 g for 10 minutes at 4°C, supernatants were transferred to new tubes and incubated for 1 hour at 4°C with 30 µl of 50% protein A agarose bead (#9863, Cell Signaling) under rotation. After a 10 minutes centrifugation at 14,000 g and 4°C, supernatants were transferred to fresh tubes and incubated with anti-RIG-I antibody under

560 rotation overnight at 4°C. Thirty microliters of protein A agarose beads were then added and
561 incubated under rotation for 30 minutes at 4°C. After centrifugation at 14,000 g for 10 minutes
562 at 4°C, pellets were washed five times with 500 µl of 1X cell lysis buffer containing 40 U/mL
563 RNasin. Pellets were resuspended with 50 µl of water and heated to 95°C for 2 minutes. After
564 centrifugation for 10 minutes at 4°C, 14,000 g RT-PCR was performed on the supernatant.

565

566 **Acknowledgements**

567 We thank Jean-Marc Cavaillon for sepsis samples and dosing IL6.

568

569

570

References

1. Jarmuz A, Chester A, Bayliss J, Gisbourne J, Dunham I, Scott J, et al. An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics*. 2002;79(3):285-96. PMID: 11863358
2. Conticello SG, Thomas CJ, Petersen-Mahrt SK, Neuberger MS. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol Biol Evol*. 2005;22(2):367-77. PMID: 15496550
3. Lecossier D, Bouchonnet F, Clavel F, Hance AJ. Hypermutation of HIV-1 DNA in the absence of the Vif protein. *Science*. 2003;300(5622):1112. PMID: 12750511
4. Mariani R, Chen D, Schrofelbauer B, Navarro F, Konig R, Bollman B, et al. Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. *Cell*. 2003;114(1):21-31. PMID: 12859895
5. Turelli P, Mangeat B, Jost S, Vianin S, Trono D. Inhibition of hepatitis B virus replication by APOBEC3G. *Science*. 2004;303(5665):1829. PMID: 15031497
6. Vartanian JP, Henry M, Marchio A, Suspène R, Aynaud MM, Guétard D, et al. Massive APOBEC3 editing of hepatitis B viral DNA in cirrhosis. *PLoS Pathog*. 2010;6(5):e1000928. PMID: 20523896
7. Bonvin M, Achermann F, Greeve I, Stroka D, Keogh A, Inderbitzin D, et al. Interferon-inducible expression of APOBEC3 editing enzymes in human hepatocytes and inhibition of hepatitis B virus replication. *Hepatology*. 2006;43(6):1364-74. PMID: 16729314
8. Argyris EG, Acheampong E, Wang F, Huang J, Chen K, Mukhtar M, et al. The interferon-induced expression of APOBEC3G in human blood-brain barrier exerts a potent intrinsic immunity to block HIV-1 entry to central nervous system. *Virology*. 2007;367(2):440-51. PMID: 17631933
9. Koning FA, Newman EN, Kim EY, Kunstman KJ, Wolinsky SM, Malim MH. Defining APOBEC3 expression patterns in human tissues and hematopoietic cell subsets. *J Virol*. 2009;83(18):9474-85. PMID: 19587057
10. Beale RC, Petersen-Mahrt SK, Watt IN, Harris RS, Rada C, Neuberger MS. Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo. *J Mol Biol*. 2004;337(3):585-96. PMID: 15019779
11. Bogerd HP, Wiegand HL, Hulme AE, Garcia-Perez JL, O'Shea KS, Moran JV, et al. Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc Natl Acad Sci USA*. 2006;103(23):8780-5. PMID: 16728505
12. Muckenfuss H, Hamdorf M, Held U, Perkovic M, Lower J, Cichutek K, et al. APOBEC3 proteins inhibit human LINE-1 retrotransposition. *J Biol Chem*. 2006;281(31):22161-72. PMID: 16735504

13. Suspène R, Aynaud M, Guétard D, Henry M, Eckhoff G, Marchio A, et al. Somatic hypermutation of human mitochondrial and nuclear DNA by APOBEC3 cytidine deaminases, a pathway for DNA catabolism. *Proc Natl Acad Sci USA*. 2011;108(12):4858-63. PMID: 21368204
14. Caval V, Suspene R, Shapira M, Vartanian JP, Wain-Hobson S. A prevalent cancer susceptibility APOBEC3A hybrid allele bearing APOBEC3B 3'UTR enhances chromosomal DNA damage. *Nature communications*. 2014;5:5129. PMID: 25298230
15. Landry S, Narvaiza I, Linfesty DC, Weitzman MD. APOBEC3A can activate the DNA damage response and cause cell-cycle arrest. *EMBO reports*. 2011;12(5):444-50. PMID: 21460793
16. Mussil B, Suspène R, Aynaud MM, Gauvrit A, Vartanian JP, Wain-Hobson S. Human APOBEC3A isoforms translocate to the nucleus and induce DNA double strand breaks leading to cell stress and death. *PLoS ONE*. 2013;8(8):e73641. PMID: 23977391
17. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature*. 2013;500(7463):415-21. PMID: 23945592
18. Burns MB, Temiz NA, Harris RS. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat Genet*. 2013;45(9):977-83. PMID: 23852168
19. Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, et al. The landscape of cancer genes and mutational processes in breast cancer. *Nature*. 2012;486(7403):400-4. PMID: 22722201
20. Komatsu A, Nagasaki K, Fujimori M, Amano J, Miki Y. Identification of novel deletion polymorphisms in breast cancer. *International journal of oncology*. 2008;33(2):261-70. PMID: 18636146
21. Zhang T, Cai J, Chang J, Yu D, Wu C, Yan T, et al. Evidence of associations of APOBEC3B gene deletion with susceptibility to persistent HBV infection and hepatocellular carcinoma. *Human molecular genetics*. 2012;22(6):1262-9. PMID: 23213177
22. Long J, Delahanty RJ, Li G, Gao YT, Lu W, Cai Q, et al. A Common Deletion in the APOBEC3 Genes and Breast Cancer Risk. *J Natl Cancer Inst*. 2013;105(8):573-9. PMID: 23411593
23. Xuan D, Li G, Cai Q, Deming-Halverson S, Shrubsole MJ, Shu XO, et al. APOBEC3 deletion polymorphism is associated with breast cancer risk among women of European ancestry. *Carcinogenesis*. 2013;34(10):2240-3. PMID: 23715497
24. Nik-Zainal S, Wedge DC, Alexandrov LB, Petljak M, Butler AP, Bolli N, et al. Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. *Nat Genet*. 2014;46(5):487-91. PMID: 24728294

25. Balkwill F, Mantovani A. Inflammation and cancer: back to Virchow? *Lancet*. 2001;357(9255):539-45. PMID: 11229684
26. Hemmi H, Takeuchi O, Kawai T, Kaisho T, Sato S, Sanjo H, et al. A Toll-like receptor recognizes bacterial DNA. *Nature*. PMID: 11130078
27. Fernandes-Alnemri T, Yu JW, Datta P, Wu J, Alnemri ES. AIM2 activates the inflammasome and cell death in response to cytoplasmic DNA. *Nature*. 2009;458(7237):509-13. PMID: 19158676
28. Oka T, Hikoso S, Yamaguchi O, Taneike M, Takeda T, Tamai T, et al. Mitochondrial DNA that escapes from autophagy causes inflammation and heart failure. *Nature*. 2012;485(7397):251-5. PMID: 22535248
29. Zhang Q, Raoof M, Chen Y, Sumi Y, Sursal T, Junger W, et al. Circulating mitochondrial DAMPs cause inflammatory responses to injury. *Nature*. 2010;464(7285):104-7. PMID: 20203610
30. Atianand MK, Fitzgerald KA. Molecular basis of DNA recognition in the immune system. *J Immunol*. 2013;190(5):1911-8. PMID: 23417527
31. Yang P, An H, Liu X, Wen M, Zheng Y, Rui Y, et al. The cytosolic nucleic acid sensor LRRFIP1 mediates the production of type I interferon via a beta-catenin-dependent pathway. *Nat Immunol*. 2010;11(6):487-94. PMID: 20453844
32. West AP, Khoury-Hanold W, Staron M, Tal MC, Pineda CM, Lang SM, et al. Mitochondrial DNA stress primes the antiviral innate immune response. *Nature*. 2015. PMID: 25642965
33. Ablasser A, Bauernfeind F, Hartmann G, Latz E, Fitzgerald KA, Hornung V. RIG-I-dependent sensing of poly(dA:dT) through the induction of an RNA polymerase III-transcribed RNA intermediate. *Nat Immunol*. 2009;10(10):1065-72. PMID: 19609254
34. Chiu YH, Macmillan JB, Chen ZJ. RNA polymerase III detects cytosolic DNA and induces type I interferons through the RIG-I pathway. *Cell*. 2009;138(3):576-91. PMID: 19631370
35. Ishikawa H, Ma Z, Barber GN. STING regulates intracellular DNA-mediated, type I interferon-dependent innate immunity. *Nature*. 2009;461(7265):788-92. PMID: 19776740
36. Zhong B, Yang Y, Li S, Wang YY, Li Y, Diao F, et al. The adaptor protein MITA links virus-sensing receptors to IRF3 transcription factor activation. *Immunity*. 2008;29(4):538-50. PMID: 18818105
37. Crow YJ, Hayward BE, Parmar R, Robins P, Leitch A, Ali M, et al. Mutations in the gene encoding the 3'-5' DNA exonuclease TREX1 cause Aicardi-Goutieres syndrome at the AGS1 locus. *Nat Genet*. 2006;38(8):917-20. PMID: 16845398
38. Yasutomo K, Horiuchi T, Kagami S, Tsukamoto H, Hashimura C, Urushihara M, et al. Mutation of DNASE1 in people with systemic lupus erythematosus. *Nat Genet*. 2001;28(4):313-4. PMID: 11479590

39. Suspène R, Henry M, Guillot S, Wain-Hobson S, Vartanian JP. Recovery of APOBEC3-edited human immunodeficiency virus G->A hypermutants by differential DNA denaturation PCR. *J Gen Virol.* 2005;86(Pt 1):125-9. PMID: 15604439
40. Aynaud MM, Suspène R, Vidalain PO, Mussil B, Guétard D, Tangy F, et al. Human Tribbles 3 protects nuclear DNA from cytidine deamination by APOBEC3A. *J Biol Chem.* 2012;287(46):39182-92. PMID: 22977230
41. Schulte W, Bernhagen J, Bucala R. Cytokines in sepsis: potent immunoregulators and potential therapeutic targets--an updated view. *Mediators of inflammation.* 2013;2013:165974. PMID: 23853427
42. Fitting C, Parlato M, Adib-Conquy M, Memain N, Philippart F, Misset B, et al. DNAemia detection by multiplex PCR and biomarkers for infection in systemic inflammatory response syndrome patients. *PLoS ONE.* 2012;7(6):e38916. PMID: 22719987
43. Zarembek KA, Godowski PJ. Tissue expression of human Toll-like receptors and differential regulation of Toll-like receptor mRNAs in leukocytes in response to microbes, their products, and cytokines. *J Immunol.* 2002;168(2):554-61. PMID: 11777946.
44. Wu L, Pan J, Thoroddsen V, Wysong DR, Blackman RK, Bulawa CE, et al. Novel small-molecule inhibitors of RNA polymerase III. *Eukaryotic cell.* 2003;2(2):256-64. PMID: 12684375
45. Suspène R, Aynaud MM, Vartanian JP, Wain-Hobson S. Efficient deamination of 5-methylcytidine and 5-substituted cytidine residues in DNA by human APOBEC3A cytidine deaminase. *PLoS ONE.* 2013;8(6):e63461. PMID: 23840298
46. Peng G, Greenwell-Wild T, Nares S, Jin W, Lei KJ, Rangel ZG, et al. Myeloid differentiation and susceptibility to HIV-1 are linked to APOBEC3 expression. *Blood.* 2007;110(1):393-400. PMID: 17371941
47. Thielen BK, McNevin JP, McElrath MJ, Hunt BV, Klein KC, Lingappa JR. Innate immune signaling induces high levels of TC-specific deaminase activity in primary monocyte-derived cells through expression of APOBEC3A isoforms. *J Biol Chem.* PMID: 20615867
48. Caval V, Suspène R, Vartanian JP, Wain-Hobson S. Orthologous mammalian APOBEC3A cytidine deaminases hypermutate nuclear DNA. *Mol Biol Evol.* 2014;31(2):330-40. PMID: 24162735
49. Crow YJ, Chase DS, Lowenstein Schmidt J, Szykiewicz M, Forte GM, Gornall HL, et al. Characterization of human disease phenotypes associated with mutations in TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, ADAR, and IFIH1. *American journal of medical genetics Part A.* 2015;167A(2):296-312. PMID: 25604658
50. Schoggins JW, Wilson SJ, Panis M, Murphy MY, Jones CT, Bieniasz P, et al. A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature.* 2011;472(7344):481-5. PMID: 21478870

- 774 51. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, et al.
775 Mutational Processes Molding the Genomes of 21 Breast Cancers. *Cell*. 2012;149(5):979-
776 93. PMID: 22608084
- 777
- 778 52. Roberts SA, Gordenin DA. Clustered and genome-wide transient mutagenesis in human
779 cancers: Hypermutation without permanent mutators or loss of fitness. *Bioessays*. 2014.
780 PMID: 24615916
- 781
- 782 53. Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, Stojanov P, et al. An
783 APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat*
784 *Genet*. 2013;45(9):970-6. PMID: 23852170
- 785
- 786 54. Denissenko MF, Chen JX, Tang MS, Pfeifer GP. Cytosine methylation determines hot
787 spots of DNA damage in the human P53 gene. *Proc Natl Acad Sci USA*. 1997;94(8):3893-
788 8. PMID: 9108075
- 789
- 790 55. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, et al. Patterns of
791 somatic mutation in human cancer genomes. *Nature*. 2007;446(7132):153-8. PMID:
792 17344846
- 793
- 794 56. Stephens PJ, McBride DJ, Lin ML, Varela I, Pleasance ED, Simpson JT, et al. Complex
795 landscapes of somatic rearrangement in human breast cancer genomes. *Nature*.
796 2009;462(7276):1005-10. PMID: 20033038
- 797
- 798 57. Taylor BJ, Nik-Zainal S, Wu YL, Stebbings LA, Raine K, Campbell PJ, et al. DNA
799 deaminases induce break-associated mutation showers with implication of APOBEC3B
800 and 3A in breast cancer kataegis. *eLife*. 2013;2:e00534. PMID: 23599896
- 801
- 802 58. Imai K, Slupphaug G, Lee WI, Revy P, Nonoyama S, Catalan N, et al. Human uracil-DNA
803 glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch
804 recombination. *Nat Immunol*. 2003;4(10):1023-8. PMID: 12958596
- 805
- 806 59. Refsland EW, Stenglein MD, Shindo K, Albin JS, Brown WL, Harris RS. Quantitative
807 profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications
808 for HIV-1 restriction. *Nucleic Acids Res*. 2010;38:4274-84. PMID: 20308164
- 809
- 810
- 811

812 **Figure Legends**

813 **Figure 1**

814 **Abundant A3-edited cymtDNA in single cells.** A) Frequencies of single cells harboring A3-
815 edited cymtDNA using a PCR denaturation temperature of 85°C. B) Number of hypo- and
816 hypermutated mtDNA sequences obtained using a PCR denaturation temperature of 95°C; ^a
817 number of analyzed sequences; ^b number of hypomutated sequences; ^c proportion of hypoedited
818 sequences per cell; ^d mean CG->TA mutation frequencies for hypoedited sequences, ^e number
819 of hyperedited mtDNA sequences identified; ^f Number (#) of C->T or G->A transitions per
820 mtDNA sequence. C) A selection of hypermutated *MT-COI* sequences. Although editing
821 occurs on both strands, the sequences are given with respect to the plus or coding strand.

823 **Figure 2**

824 ***MT-COI* editing in P2 single cell.** A) 3DPCR recovered edited *MT-COI* DNA genomes down
825 to 86.1°C for single cells B02 and down to 84.6°C for single cell E11. The white line indicates
826 the threshold between edited and unedited 3DPCR products. Asterisks refer to the samples
827 cloned and sequenced. Cell D05 did not harbor edited mtDNA and used as control. M:
828 molecular weight markers. B) Mutation matrices for hyperedited *MT-COI* DNA sequences
829 from cells D05, B02 and E11. n indicates the number of bases sequenced. C) Dinucleotide
830 context of *MT-COI* DNA region minus strand DNA obtained in cell B02 and E11. Chi-square
831 test indicates dinucleotide frequencies that significantly deviate from expected values (*
832 p<0.05). D) Clonal analysis of *MT-COI* editing for individual edited sequences from cell B02
833 and E11. Blue asterisks represent the overlapping sequences between cell B02 and E11. The
834 number of TpC+CpC vs. GpC+ApC targets edited per sequence and represented on the y and x
835 axes respectively (right), and clonal analysis using TpC vs. CpC (left). E) Edited *MT-COI* DNA
836 from P2 cell B02 at a single cell level in two different dinucleotide context, GpA (clone 1) and
837 GpG (clone 2). F) Bulk dinucleotide context of quail *MT-COI* DNA edited by A3A, A3C, A3F
838 and A3G cytidine deaminases, (* p<0.05). G) Clonal analysis of *MT-COI* editing for A3A,
839 A3C, A3F and A3G cytidine deaminases.

841 **Figure 3**

842 **Etoposide and actinomycin D induced A3 expression and *MT-COI* editing in P2 cells.** A)
843 Transcription profiling of A3A-A3H in etoposide or actinomycin D treated-P2 cells. Data in
844 duplicate were normalized to the expression levels of *RPL13A* housekeeping reference genes
845 and to untreated P2 cells to facilitate comparison. B, C) FACS analysis of cytochrome c release

(A) and apoptosis (B) in P2 cells treated with 100 μ M of actinomycin D or 100 μ M etoposide after 16 hours. Annexin V scored early apoptosis and propidium iodide (PI) late apoptosis/necrosis. Means and SEM are given for two independent experiments. D) 3DPCR recovered edited *MT-COI* DNA genomes down to 83.8-84.6°C for single cells A04 and F06 treated with actinomycin D and for P2 single cells D08 and E10 treated with etoposide. Cell C08 served as unedited control. The white line indicates the threshold between edited and unedited 3DPCR products in terms of the denaturation temperature. Asterisks refer to the samples cloned and sequenced. M: molecular weight markers. E) Dinucleotide context for A3-edited *MT-COI* DNA. Chi-square test analysis indicates dinucleotide frequencies that significantly deviate from the expected values (* $p < 0.05$). F) Mutation matrices for hyperedited *MT-COI* DNA sequences derived from cloned 3DPCR products. n indicates the number of bases sequenced.

Figure 4

***MT-COI* editing in healthy and sepsis patients.** A) Schematic representing the denaturation temperature of the last positive 3DPCR amplification for *MT-COI* DNA derived from the serum of sepsis and healthy patients. Black circle DNA indicates a molecular *COI-MT* DNA clone of the reference sequence, white circles represent serum samples from sepsis patients and grey circles represent serum samples from healthy patients (* Fisher analysis, $p = 0.024$). B) Sequence analysis of hyperedited sequences obtained from sepsis patients (S) and healthy controls (H). ^a Amount of serum IL6 (pg/mole), ^b % of G->A edits, ^c YpC/RpC was calculated as follows: $((\text{TpC} + \text{CpC}) / (\text{GpC} + \text{ApC}))_{\text{observed}} / ((\text{TpC} + \text{CpC}) / (\text{GpC} + \text{ApC}))_{\text{expected}}$. A value > 1 is indicative of A3 cytidine deamination in TpC+CpC dinucleotide context. nd: not determined.

Figure 5

Uracil containing DNA is not a novel danger signal. A) Agarose gel of *MT-COI* PCR products amplified with *Taq* or *Pfu* polymerase in the presence of dTTP, dUTP, dTTP+dUTP. M, molecular weight markers. B, C) Interferon α and β production following transfection of THP-1 cells by dT or dU containing PCR DNA fragments. *MT-COI** indicates incubation with DNA but no transfection. D) *APOBEC3* transcription profiling of transfected THP-1 cells. Data were normalized to the expression levels of *RPL13A* housekeeping reference gene. Profiling was performed in duplicate and normalized to JetPRIME to facilitate comparison. E) THP-1 DNA transfection increases A3-editing of endogenous cytoplasmic *MT-CYB* DNA recovered

880 by 3DPCR. The white line indicates the threshold between edited and unedited 3DPCR
881 products. Asterisks refer to the samples cloned and sequenced. M: molecular weight markers.
882 F) Frequency distribution of G->A editing per clone as a function of the quantity of DNA
883 transfected.

884

885 **Figure 6**

886 **Transfected DNA induces double-stranded DNA breaks.** A) Interferon α production
887 following transfection of THP-1 cells by 500 ng *MT-COI_T* or *MT-COI_U*, 500 ng of WT₁₊₂,
888 WT₂₊₃, Hyp₁₊₂, Hyp₃₊₄, WT₁₊₃ and WT₂₊₄ (see Table S1). B) FACS analysis of γ H2AX-positive
889 THP-1 cells transfected with *MT-COI*. Means and standard deviation of γ H2AX-positive cell
890 frequencies for duplicate transfection. C) FACS analysis of γ H2AX-positive DSBs in THP-1
891 transfected with WT_{3+2FAM}, WT_{3+2FAM}, W_{1+2FAM} and W_{1+2FAM} respectively. Means and standard
892 deviation of γ H2AX-positive cell frequencies for duplicate transfection. D) FACS analysis of
893 γ H2AX-positive DSBs in THP-1 transfected with cells gated on FAM-positive cells after
894 transfection with WT_{3+2FAM} and W_{1+2FAM} respectively. Means and standard deviation of
895 γ H2AX-positive cell frequencies for duplicate transfection. E) Transcription profiling of A3A
896 in THP-1 transfected cells with WT_{1+2FAM} (dsDNA) or WT_{3+2FAM} (ssDNA) respectively, in
897 presence or absence of JetPRIME. * indicates statistically significant difference between two
898 observed percentages ($p < 0.05$).

899

900 **Figure 7**

901 **Transfected dsDNA triggers the RIG-I pathway via RNA polymerase III.** A) Western blots
902 of TBK1, TBK1-P (phosphorylated), IRF3, IRF3-P (phosphorylated), RIG-I, MDA5, MAVS,
903 STING, IKK ϵ at 24 hours post DNA transfection of THP-1 cells. B) Interferon alpha
904 production by DNA transfected THP-1 cells along with RNA Polymerase III inhibitor (ML-
905 60218). Peptidoglycans (PGN) were used as control. C) A3-edited *MT-CYB* DNA recovered by
906 3DPCR in presence or absence of 25-75 μ M of RNA polymerase III inhibitor. The white line
907 indicates the threshold between edited and unedited 3DPCR products. M: molecular weight
908 markers. D) Interferon alpha and beta production by THP-1 following transfection by 500 ng
909 V1V2 HIV-1 DNA. JetPRIME was used as control. E) V1V2 RNA transcripts from DNA
910 transfected THP-1 cells. Total RNA was extracted and a cDNA corresponding to V1V2 was
911 produced in presence or absence of RT. T1 and T2 refer to independent transfections. F)
912 Immunoprecipitation with an anti-RIG-I monoclonal antibody performed along with an anti-

HA as control. V1V2 specific RT-PCR products were recovered only from anti-RIG-I immunoprecipitation. RT, T1 and T2 as for 7E.

Figure 8

Cytosolic DNA mediated innate immune signaling. Upon cellular stress (1), mtDNA is released into the cytoplasm (2) is recognized by RNA polymerase III and transcribed into RNA which anneals to form dsRNA duplexes (3). These activate RIG-I signaling leading to IRF-3 phosphorylated by TBK1 (4-5) and induction of interferon (6). Through IFNAR1/2, signaling IFN production leads to A3A up-regulation (7), which initiates catabolism of cytoplasmic ssDNA. As A3A can translocate to the nucleus, it can cause hypermutation of nuDNA and formation of double stranded DNA breaks DSB (8).

947 **Supporting Information**

948 **S1 Figure**

949 **Editing of *MT-COI* in single CD4⁺ T lymphocytes from donor 1.** A) 3DPCR recovered
950 edited *MT-COI* DNA genomes down to 83-83.8°C for single cell A06 and F09. The white line
951 indicates the threshold between edited and unedited 3DPCR products in terms of the
952 denaturation temperature. Cell B05 was not edited and was used as control. Asterisks refer to
953 the samples cloned and sequenced. M: molecular weight markers. B) Mutation matrixes for
954 hyperedited *MT-COI* DNA sequences from single cells A06 and F09 (Td=86.1°C). n indicates
955 the number of bases sequenced. C) Dinucleotide context of *MT-COI* DNA region minus strand
956 DNA obtained from cells A06 and F09 at Td=86.1°C. Chi-square test indicates dinucleotide
957 frequencies that significantly deviate from the expected values (* p<0.05). D) TaqMan analysis
958 of positively selected CD4⁺ T cells from donor 1. Data in duplicate were normalized to the
959 expression levels of *RPL13A* housekeeping reference genes and normalized to untreated P2
960 cells to facilitate comparison.

961

962 **S2 Figure**

963 **Editing of *MT-COI* in single CD4⁺ T lymphocytes from donor 2.** A) 3DPCR recovered
964 edited *MT-COI* DNA genomes down to 82.2-85.4°C for single cell E02 and C08. The white
965 line indicates the threshold between edited and unedited 3DPCR products in terms of the
966 denaturation temperature. Cell G07 was not edited and was used as control. Asterisks refer to
967 the samples cloned and sequenced. M: molecular weight markers. B) Mutation matrixes for
968 hyperedited *MT-COI* DNA sequences from single cells E02 and C08 (Td=86.1°C). n indicates
969 the number of bases sequenced. C) Dinucleotide context of *MT-COI* DNA region minus strand
970 DNA obtained from cells E02 and C08 at Td=86.1°C. Chi-square test analysis indicates
971 dinucleotide frequencies that significantly deviate from the expected values (* p<0.05). D)
972 TaqMan analysis of positively selected CD4⁺ T cells from donor 2. Data in duplicate were
973 normalized to the expression levels of *RPL13A* housekeeping reference genes and normalized
974 to untreated P2 cells to facilitate comparison.

975

976 **S3 Figure**

977 **A3 transcriptome of DNA transfected THP-1 cells.** Transcriptomes were established for
978 DNA containing dU (A) or a 50:50 mixture of dT and dU (B). Duplicates were normalized to
979 the expression levels of *RPL13A* housekeeping gene and normalized to untreated THP-1 cell to
980 facilitate comparison.

981 **S4 Figure**

982 **Transfected *MT-COI* DNA activates IFN α and IFN β in PBMCs.** Interferon α (A) and
983 interferon β (B) production following treatment by a variety of DNAs as well as polyIC (500
984 ng/ml) and CpG (500 ng/ml). *MT-COI** indicates incubation with DNA without transfection.
985 C) Transcription profiling of *A3A* in THP-1 in presence or absence of IFN α for 16h or 24h.
986 Controls were performed with 500 ng *MT-COI_T* or jetPRIME only for 24h.

987

988 **S1 Table**

989 **Compendium of oligonucleotides used in this study.**

990

991

992

993

994

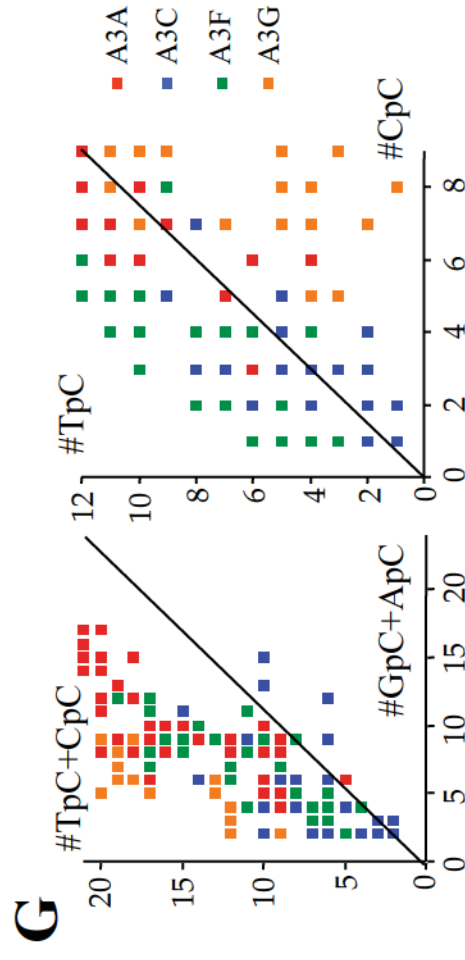
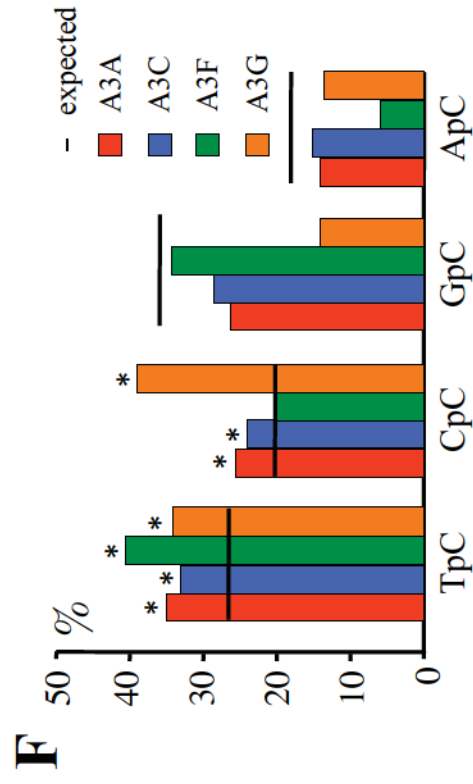
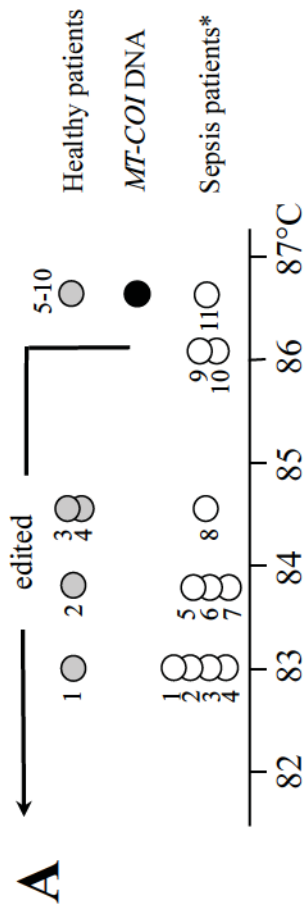


Figure 2-2



B

Patients	IL6 ^a	%CG->TA ^b	YpC/RpC ^c
S01	1973	54	3.23
S02	621	64	3.91
S03	15181	49	2.77
S04	2135	49	2.77
S05	2356	59	2.34
S06	3200	73	4.46
S07	144	59	3.39
S08	677	57	2.16
S09	2315	35	1.76
S10	14507	45	2.35
H01	nd	58	3.49
H02	nd	53	3.80
H03	nd	59	3.23
H04	nd	36	1.38

Figure 3

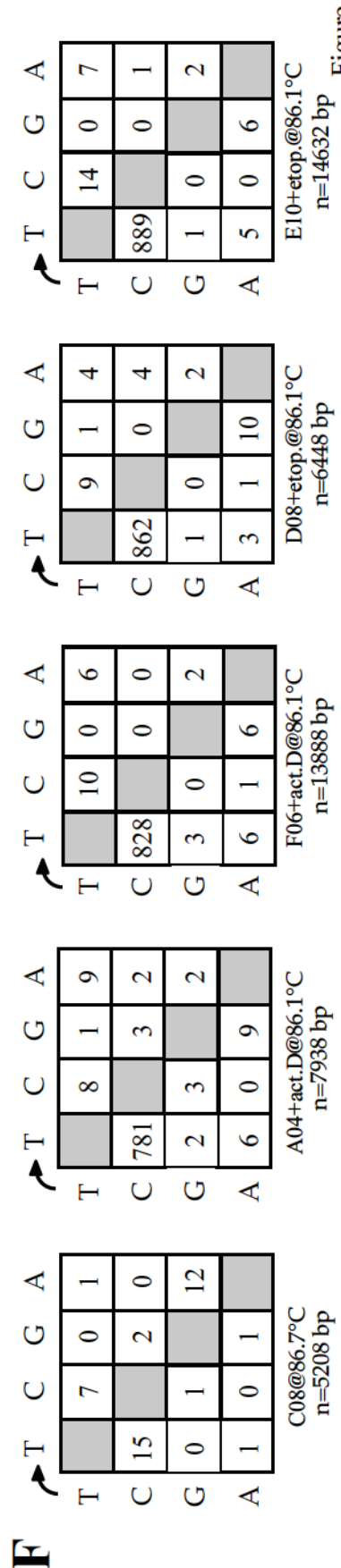
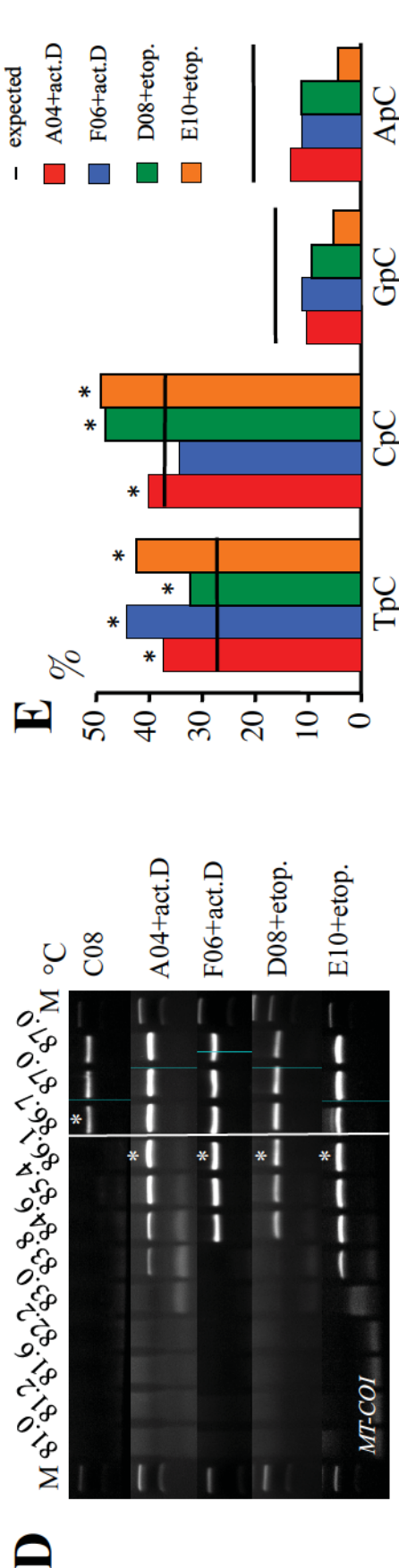
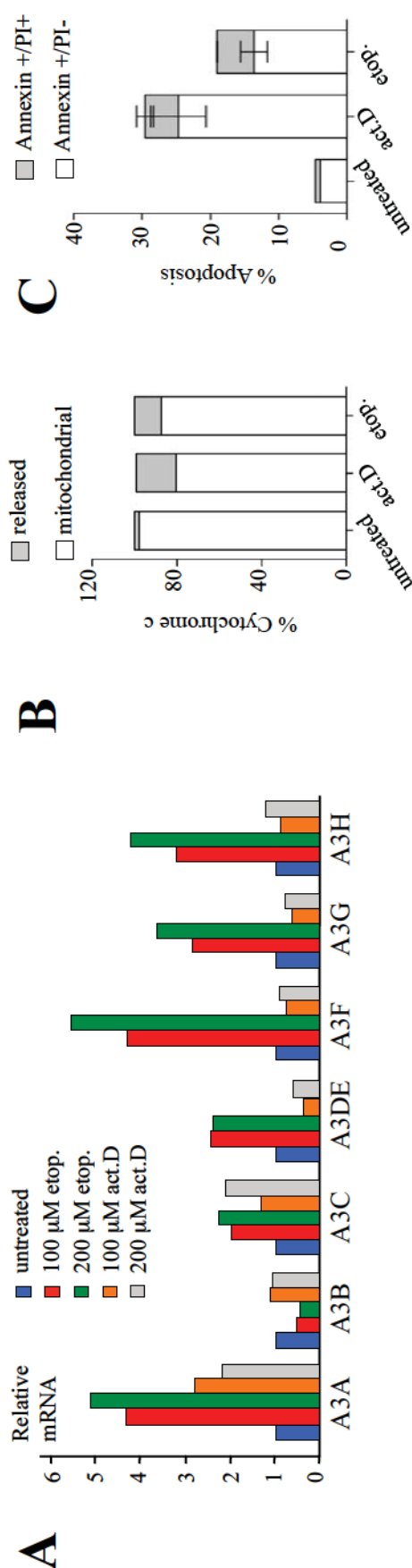


Figure 4

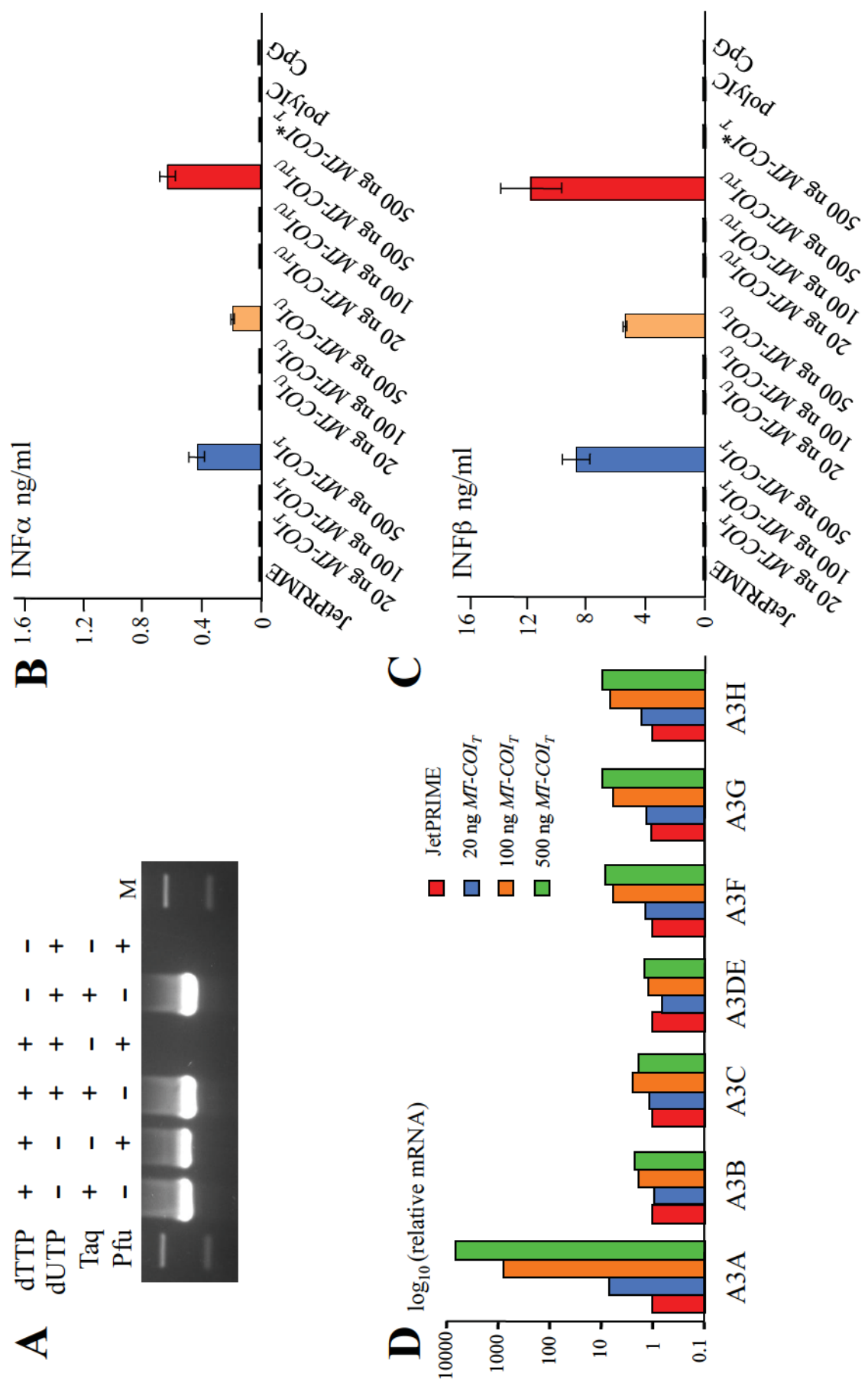


Figure 5-1

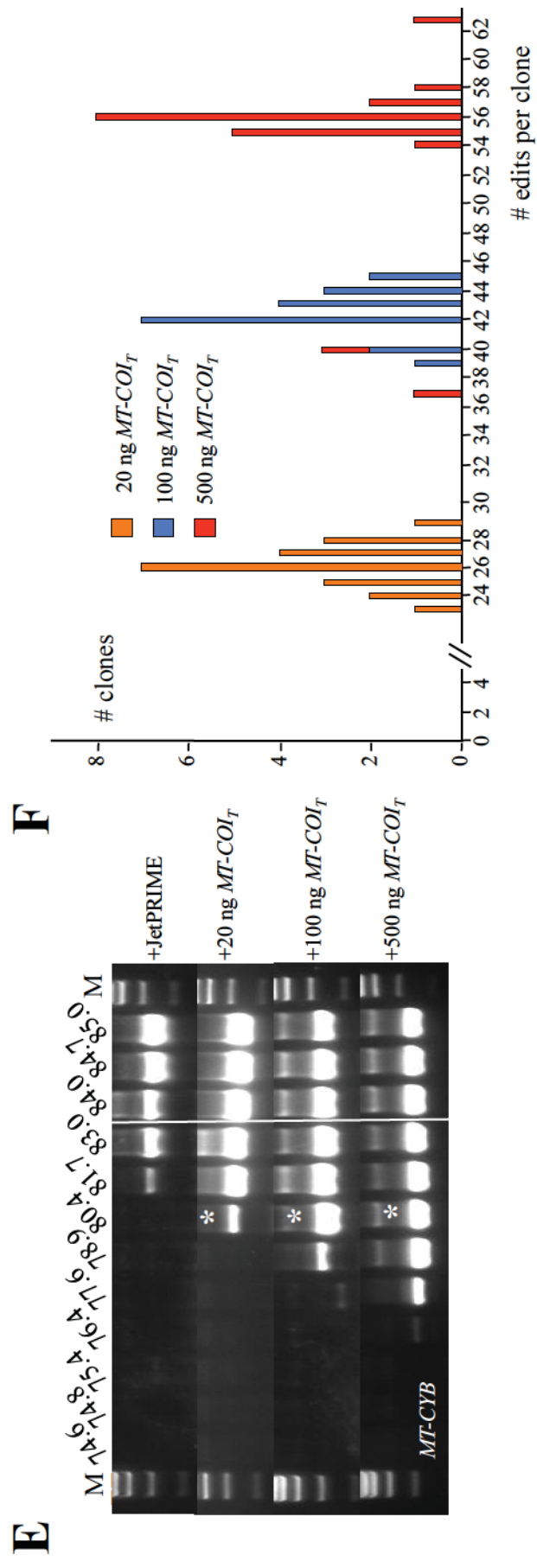


Figure 5-2

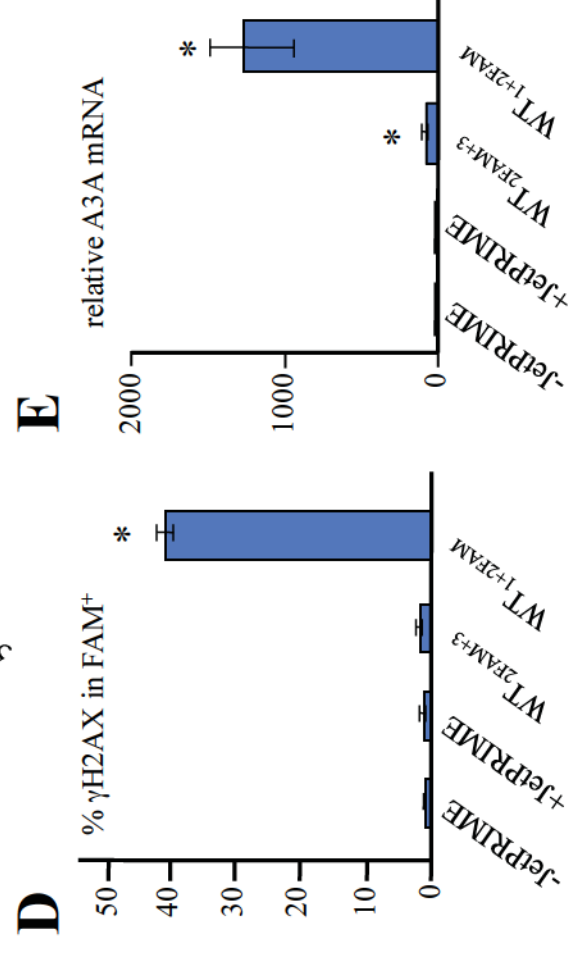
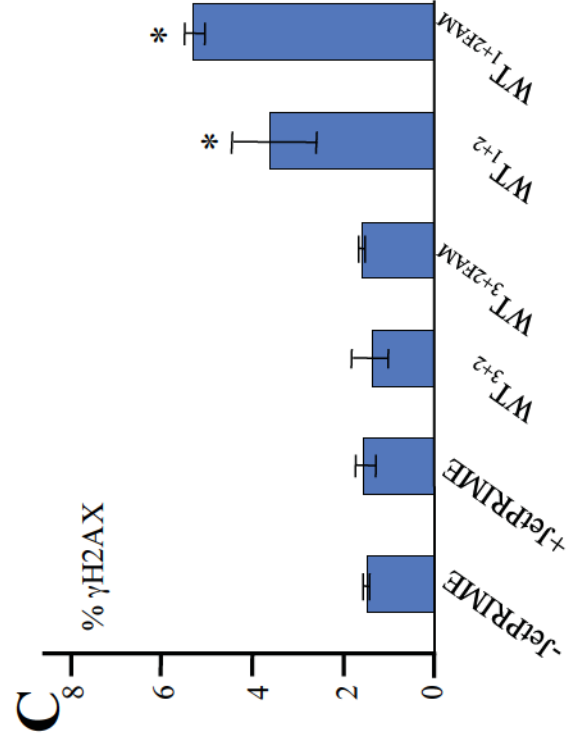
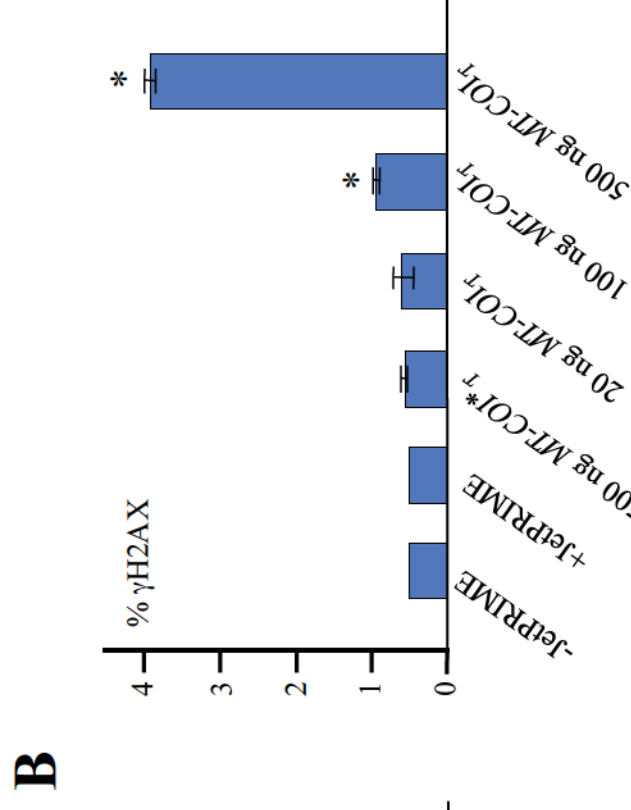
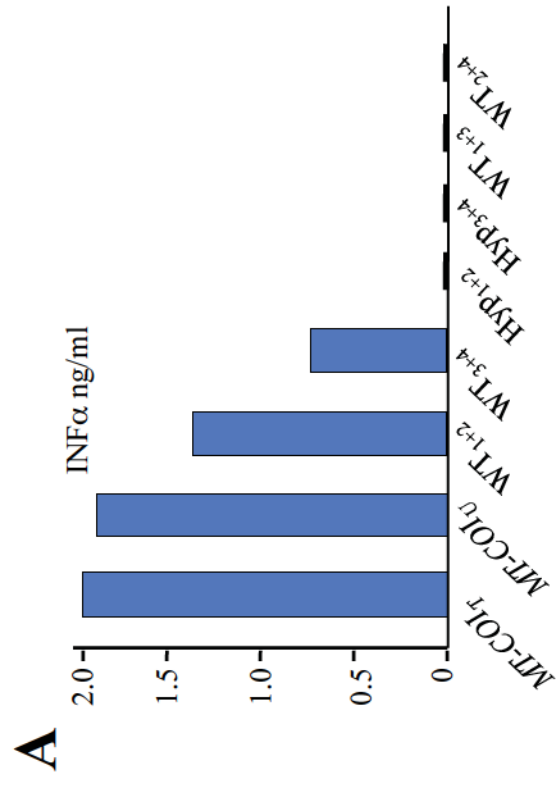


Figure 6

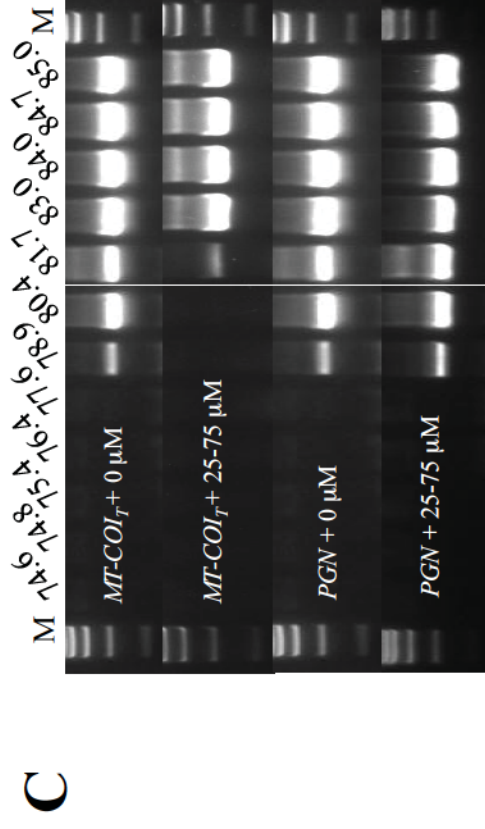
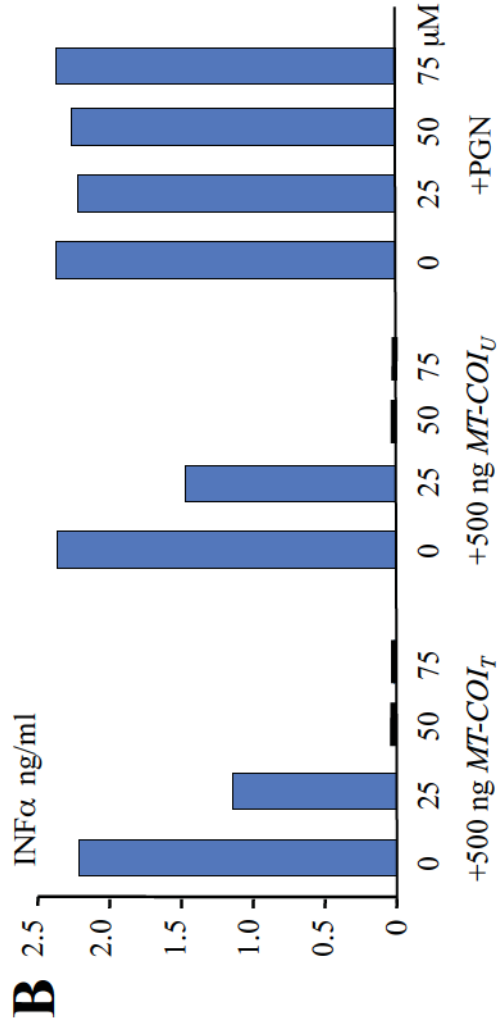
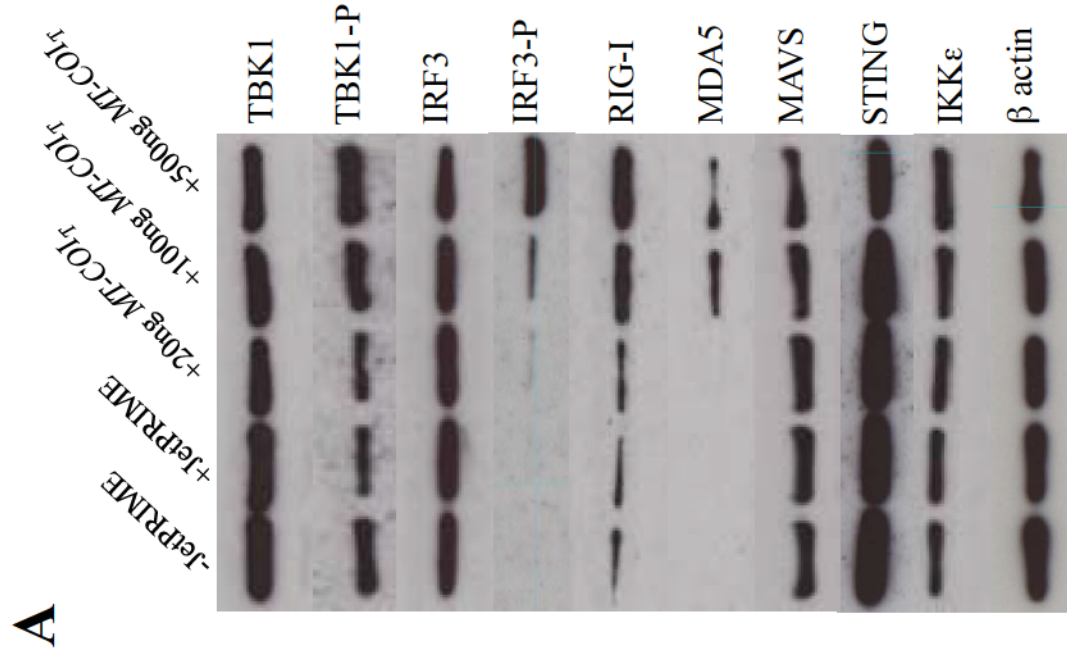


Figure 7-1

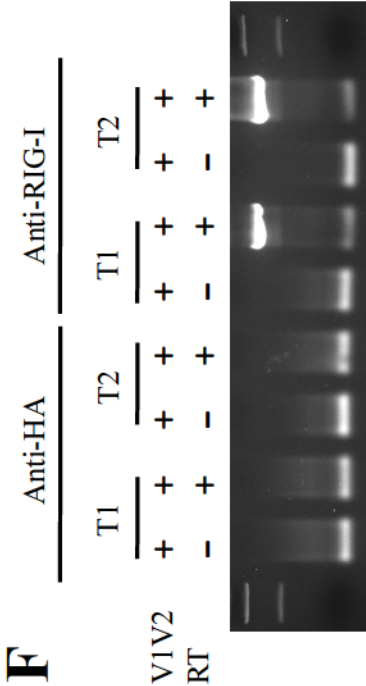
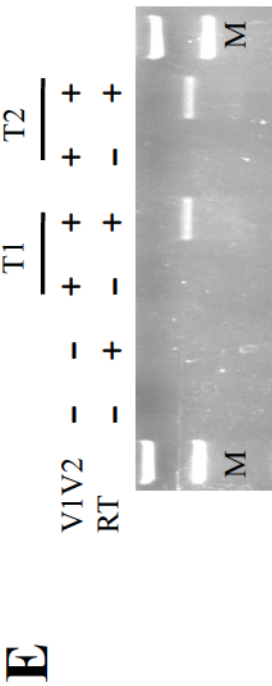
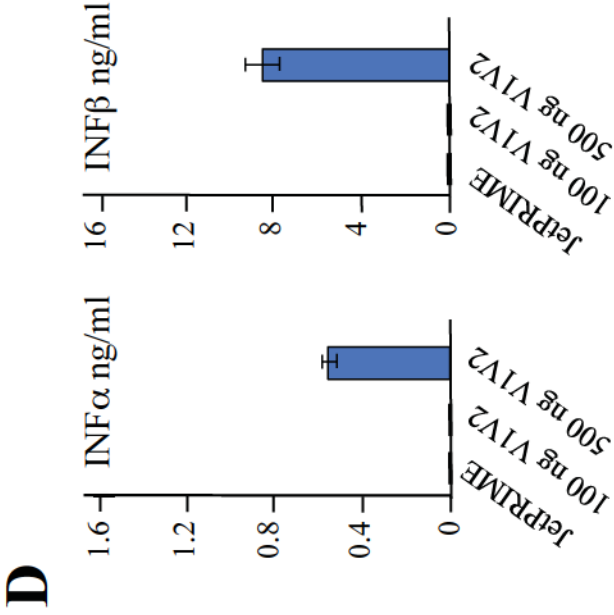


Figure 7-2

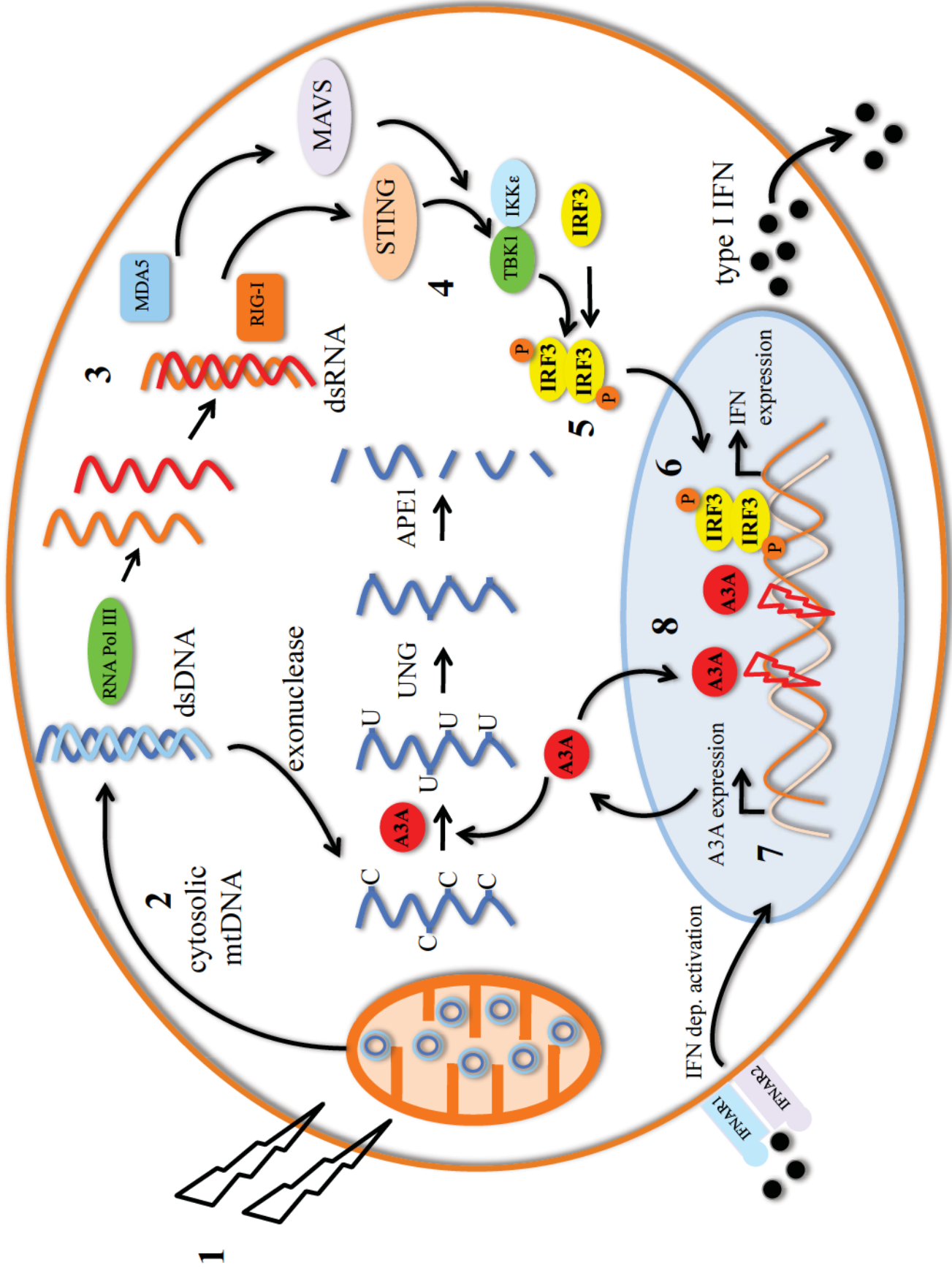


Figure 8

Figure S1

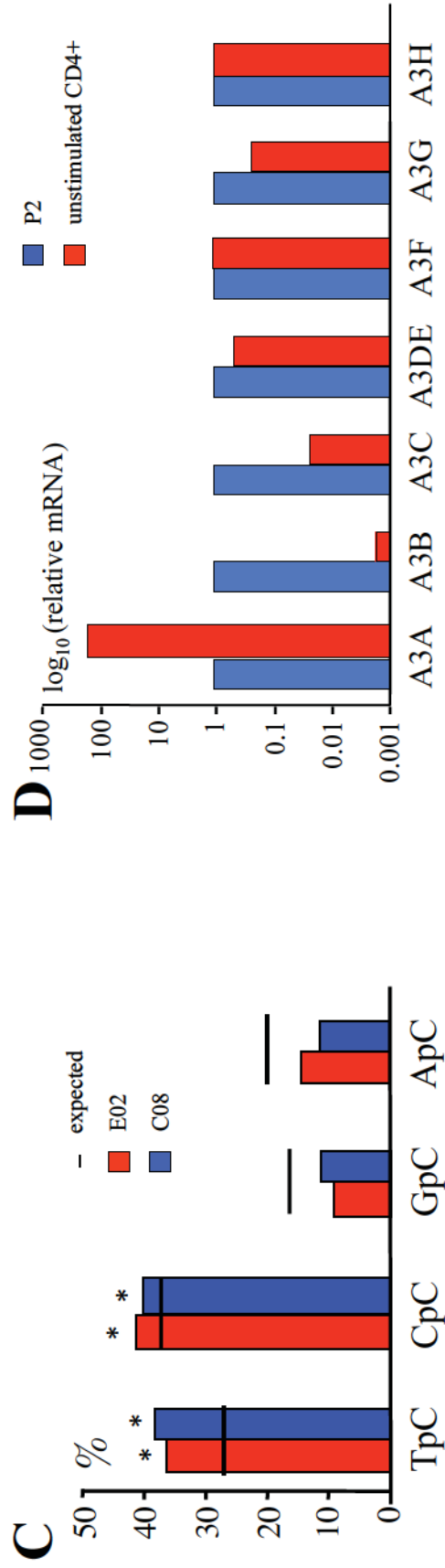
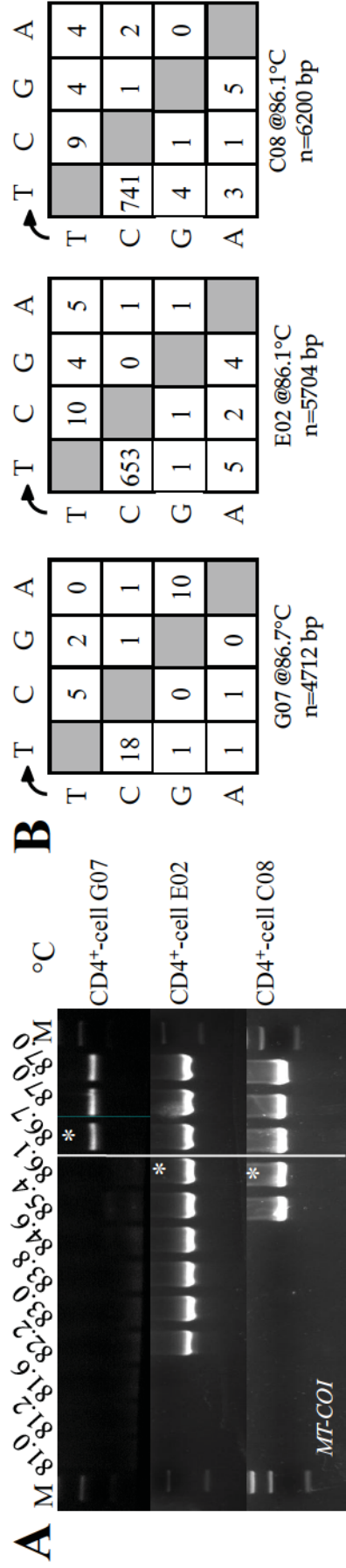


Figure S2

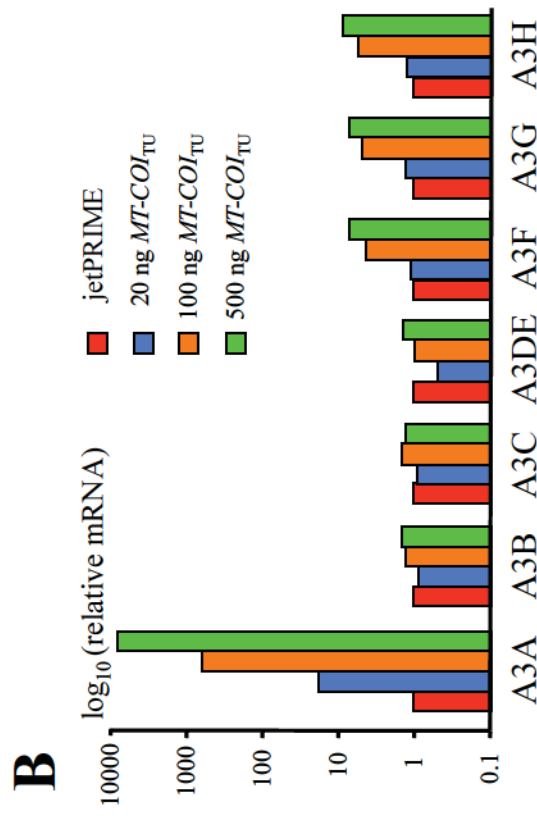
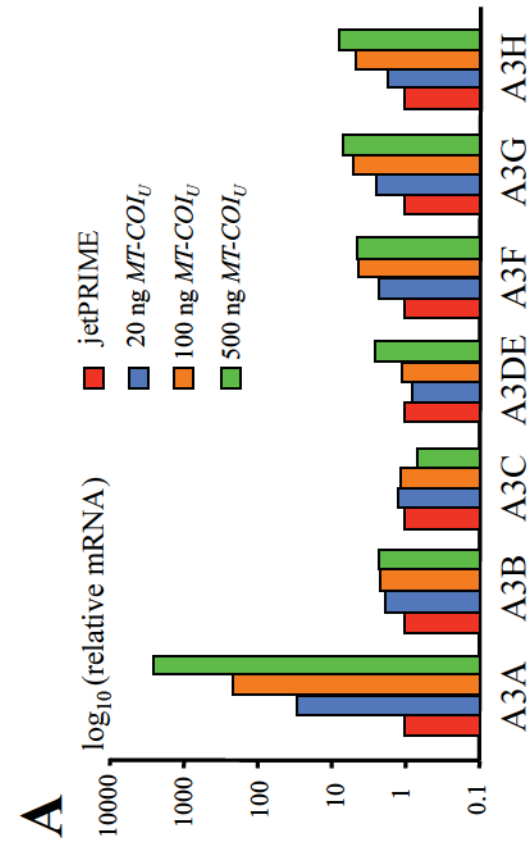


Figure S3

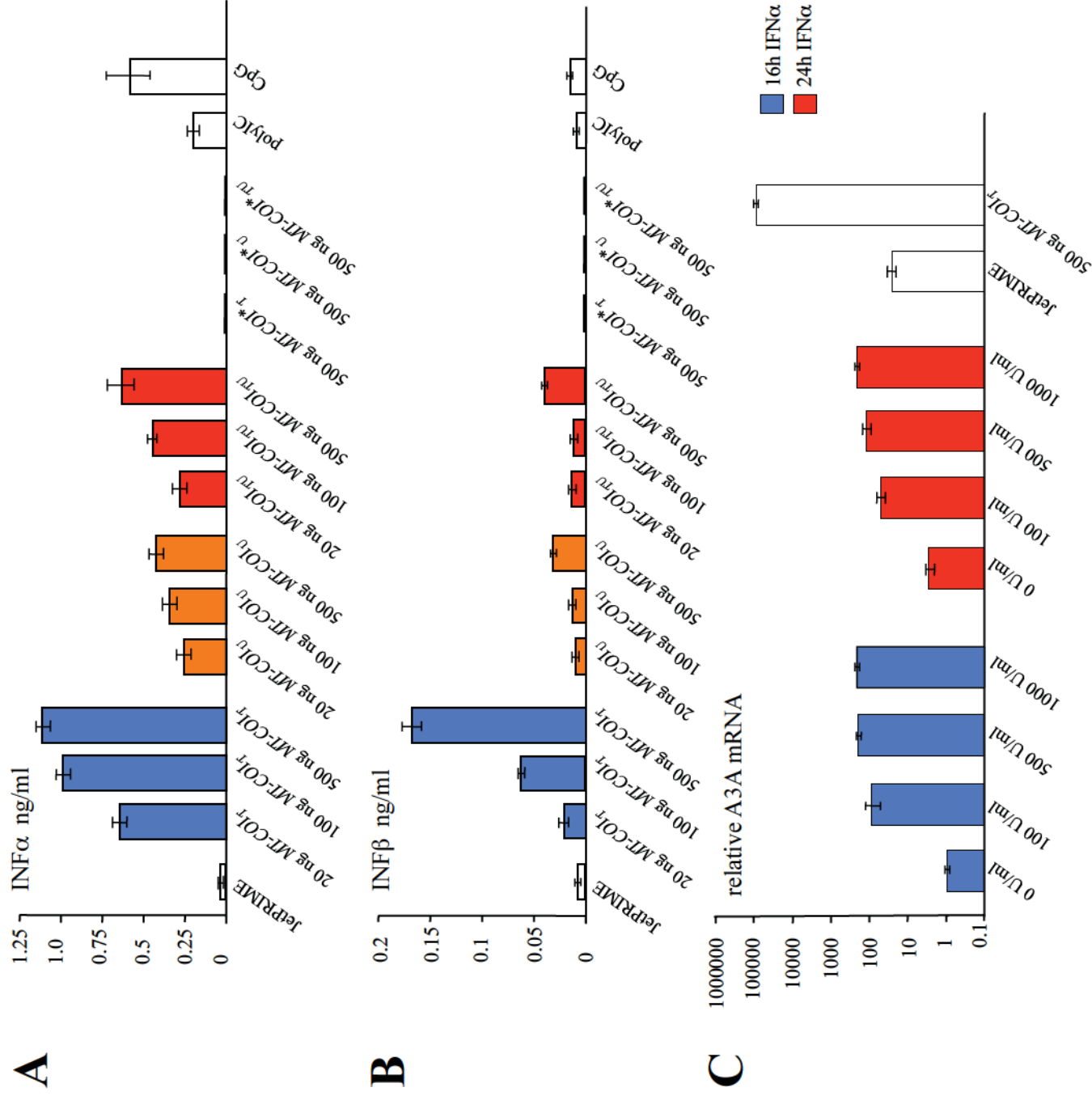


Figure S4

Primer	Sequence
WT ₁	5 ' ACTCTTACCTCCCTCTCTCCTACTCCTGCTCGCATCTGCTATA
WT ₂	5 ' TATAGCAGATGCGAGCAGGAGTAGGAGAGAGGGAGGTAAGAGT
Hyp ₁	5 ' ATTTTATTTTTTTTTTTTTTTTATTTTGTGTTGTATTTGTTATA
Hyp ₂	5 ' TATAACAAATACAAACAAAAATAAAAAAAAAAAAAATAAAAAAT
WT ₃	5 ' AACTACTCCCACCCTGGAGCCTCCGTAGACCTAACCATC
WT ₄	5 ' GATGGTTAGGTCTACGGAGGCTCCAGGGTGGGAGTAGTT
Hyp ₃	5 ' AATTATTTTTATTTTGGAGTTTTTGTAGATTTAATTATT
Hyp ₄	5 ' AATAATTAAATCTACAAAACTCCAAAATAAAAAATAGTT

S1 Table

Résumé

Les protéines APOBEC3 (A3A-A3H) catalysent la désamination des cytidines (C) présentes sur l'ADN simple brin en thymidine (T).

Cette activité cytidines désaminase a initialement été décrite comme impliquée dans la restriction des rétrovirus et de certains virus à ADN par leur capacité à induire de nombreuses mutations C->T, ou hypermutations, sur les génomes viraux. Il apparaît néanmoins que leur activité n'est pas restreinte aux génomes viraux et que certaines A3 peuvent induire des mutations sur l'ADN mitochondrial (A3A, C, F, G et H) et nucléaire (A3A et A3B). Ainsi, l'impact somatique des A3 est désormais établi dans la formation de certains cancers, dont la majorité des mutations, portent signatures des APOBEC3.

Aux vues de ces observations, nous nous sommes intéressés à la façon dont sont régulées ces enzymes dans le contexte du stress cellulaire viro-induit ou endogène.

La première partie de nos travaux a porté sur la protéine A3DE, seul membre de la famille APOBEC3 ne possédant pas d'activité cytidine désaminase. De façon intéressante, il apparaît qu'A3DE est surexprimée dans les cirrhoses infectées par le VHB, VHC ou co-infectées par le VHC et le VHB. Nous avons pu mettre en évidence qu'A3DE interagit et module l'activité d'A3F et d'A3G, deux cytidines désaminases exprimées dans le foie et impliquées dans la restriction du VHB.

Dans un second temps, nous nous sommes intéressés à la caractérisation du potentiel génotoxique de la protéine A3B. Cette protéine, de par sa localisation strictement nucléaire, constitue la seule A3 à double domaine n'interagissant pas avec A3DE. Contrairement à A3A, A3B est faiblement active sur l'ADN nucléaire et n'induit pas de cassures de l'ADN double brin. Nous avons pu mettre en évidence par mutagenèse les régions de la protéine impliquées dans l'atténuation de la génotoxicité d'A3B par rapport à A3A et que cette atténuation est conservée chez les primates. Enfin, nous avons étudié le rôle et la régulation d'A3A dans le catabolisme. Nous avons mis en évidence que l'ADN mitochondrial cytoplasmique (ADNcymt) active la voie RIG-I/ARN polymérase III ce qui a pour effet d'induire la production d'IFN qui va activer l'expression d'A3A. A3A va ainsi jouer un rôle dans le catabolisme de l'ADNcymt et contribue à l'élimination de cette source de stress cellulaire, mais occasionnant par la même des dommages sur l'ADN nucléaire.

Les A3 sont des enzymes fondamentales de la défense immunitaire innée et du catabolisme de l'ADN. Nous montrons qu'A3DE a pour fonction de moduler l'activité d'A3F et d'A3G tandis qu'A3B, possède un phénotype atténué chez tous les primates et s'avère moins génotoxique que'A3A. Cette dernière participe à la dégradation de l'ADN cytoplasmique, limitant ainsi l'inflammation. Néanmoins, A3A peut s'avérer dangereuse pour l'intégrité génomique et contribuer à l'émergence de cancers, notamment en cas d'inflammation chronique.